Rolly Intan
Chi-Hung Chi
Henry N. Palit
Leo W. Santoso (Eds.)

Communications in Computer and Information Science 516

Intelligence in the Era of Big Data

4th International Conference on Soft Computing, Intelligent Systems and Information Technology, ICSIIT 2015 Bali, Indonesia, March 11–14, 2015, Proceedings

Springer

Communications in Computer and Information Science 516

More information about this series at http://www.springer.com/series/7899

Rolly Intan · Chi-Hung Chi
Henry N. Palit · Leo W. Santoso (Eds.)

Intelligence in the Era of Big Data

4th International Conference
on Soft Computing, Intelligent Systems
and Information Technology, ICSIIT 2015
Bali, Indonesia, March 11–14, 2015
Proceedings

Springer

Editors
Rolly Intan
Informatics
Petra Christian University
Surabaya
Indonesia

Henry N. Palit
Informatics
Petra Christian University
Surabaya
Indonesia

Chi-Hung Chi
CSIRO
Hobart
Tasmania
Australia

Leo W. Santoso
Informatics
Petra Christian University
Surabaya
Indonesia

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-662-46741-1 ISBN 978-3-662-46742-8 (eBook)
DOI 10.1007/978-3-662-46742-8

Library of Congress Control Number: 2015934823

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media (www.springer.com)

Preface

This proceedings volume contains papers presented at the fourth International Conference on Soft Computing, Intelligent System and Information Technology (the 4th ICSIIT) held in Bali, Indonesia, during March 11–14, 2015. The main theme of this international conference is "Intelligence in the Era of Big Data," and it was organized and hosted by Informatics Engineering Department, Petra Christian University, Surabaya, Indonesia.

The Program Committee received 92 submissions for the conference from across Indonesia and around the world. After peer-review process by at least two reviewers per paper, 53 papers were accepted and included in the proceedings. The papers were divided into 14 groups: fuzzy logic and control system, genetic algorithm and heuristic approaches, artificial intelligence and machine learning, similarity-based models, classification and clustering techniques, intelligent data processing, feature extraction, image recognition, visualization technique, intelligent network, cloud and parallel computing, strategic planning, intelligent applications, and intelligent systems for enterprise government and society.

We would like to thank all Program Committee members for their effort in providing high-quality reviews in a timely manner. We thank all the authors of submitted papers and the authors of selected papers for their collaboration in preparation of the final copy.

Compared to the previous ICSIIT conferences, the number of participants at the 4th ICSIIT 2015 is not only higher, but also the research papers presented at the conference are improved both in quantity and quality. On behalf of the Organizing Committee, once again, we would like to thank all the participants of this conference, who contributed enormously to the success of the conference.

We hope all of you enjoy reading this volume and that you will find it inspiring and stimulating for your research and future work.

February 2015

Rolly Intan
Chi-Hung Chi
Henry N. Palit
Leo W. Santoso

Organization

The International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT) 2015 (http://icsiit.petra.ac.id) took place in Bali, Indonesia, during March 11–14, 2015, hosted by Informatics Department, Petra Christian University.

General Chair

Leo Willyanto Santoso	Petra Christian University, Indonesia

Program Chairs

Chen Ding	Ryerson University, Canada
Justinus Andjarwirawan	Petra Christian University, Indonesia
Wei Zhou	CSIRO, Australia

Registration Chairs

Silvia Rostianingsih	Petra Christian University, Indonesia

Local Arrangement Chairs

Agustinus Noertjahyana	Petra Christian University, Indonesia

Financial Chairs

Alexander Setiawan	Petra Christian University, Indonesia

Program Committee

A. Min Tjoa	Vienna University of Technology, Austria
A.V. Senthil Kumar	Hindusthan College of Arts and Science, India
Achmad Nizar Hidayanto	University of Indonesia, Indonesia
Alexander Fridman	Institute for Informatics and Mathematical Modelling, Russia
Arif Anjum	University of Pune, India
Ashraf Elnagar	University of Sharjah, United Arab Emirates
Bruce Spencer	University of New Brunswick, Canada
Byung-Gook Lee	Dongseo University, Korea

Can Wang	CSIRO, Australia
Chi-Hung Chi	CSIRO, Australia
Dengwang Li	Shandong Normal University, China
Eduard Babulak	Maharishi University of Management in Fairfield, USA
Enrique Dominguez	University of Malaga, Spain
Erma Suryani	Sepuluh Nopember Institute of Technology, Indonesia
Felix Pasila	Petra Christian University, Indonesia
Hans Dulimarta	Grand Valley State University, USA
Henry N. Palit	Petra Christian University, Indonesia
Hong Xie	Murdoch University, Australia
Ibrahiem M. M. El Emary	King Abdulaziz University, Saudi Arabia
Ilung Pranata	The University of Newcastle, Australia
Julien Dubois	Université de Bourgogne, France
Kassim S. Mwitondi	Sheffield Hallam University, UK
Kelvin Cheng	National University of Singapore, Singapore
Marian S. Stachowicz	University of Minnesota, USA
Masashi Emoto	Meiji University, Japan
Mehmed Kantardzic	University of Louisville, USA
Moeljono Widjaja	Agency for the Assessment and Application of Technology, Indonesia
Mohd Yunus Bin Nayan	Universiti Teknologi Petronas, Malaysia
Muhammad Aamir Cheema	Monash University, Australia
Noboru Takagi	Toyama Prefectural University, Japan
Nur Iriawan	Sepuluh Nopember Institute of Technology, Indonesia
P.S. Avadhani	Andhra University, India
Pitoyo Hartono	Chukyo University, Japan
Pujianto Yugopuspito	Pelita Harapan University, Indonesia
Raymond Kosala	Binus University, Indonesia
Raymond Wong	University of New South Wales, Australia
Roberto Rojas-Cessa	New Jersey Institute of Technology, USA
Rolly Intan	Petra Christian University, Indonesia
Rudy Setiono	National University of Singapore, Singapore
S. Thabasu Kannan	Pannai College of Engineering and Technology, India
Sankar Kumar Pal	Indian Statistical Institute, India
Saurabh K. Garg	University of Tasmania, Australia
Selpi	Chalmers University of Technology, Sweden
Shafiq Alam Burki	University of Auckland, New Zealand
Shan-Ling Pan	University of New South Wales, Australia
Simon Fong	University of Macau, Macau
Smarajit Bose	Indian Statistical Institute, India

Son Kuswadi	Electronic Engineering Polytechnic Institute of Surabaya, Indonesia
Suash Deb	CV Raman College of Engineering, India
Suphamit Chittayasothorn	King Mongkut's Institute of Technology Ladkrabang, Thailand
Taweesak Kijkanjanarat	Thammasat University, Thailand
Vatcharaporn Esichaikul	Asian Institute of Technology, Thailand
Vincent Vajnovszki	Université de Bourgogne, France
Wen-June Wang	National Central University, Taiwan
Wichian Chutimaskul	King Mongkut's University of Technology Thonburi, Thailand
Xiaojun Ye	Tsinghua University, China
Yung-Chen Hung	Soochow University, Taiwan
Yunwei Zhao	Tsinghua University, China

Keynote and Invited Papers

Data Mining Model for Road Accident Prediction in Developing Countries

Sanjay Misra

Covenant University, Canaanland, Ogun State, Ota, Nigeria
sanjay.misra@covenantuniversity.edu.ng

Abstract. Human loss due to road traffic accident (RTA) in developing countries is a big challenge. It becomes more serious in those developing countries where road conditions are not good and due to several reasons government is not able to maintain roads on regular basis. Additionally, increasing number of vehicles, inefficient driving and environmental conditions are also some of the factors which are responsible for RTA. In this work we present architecture of a data mining model. The proposed model is applied on real data set of RTAs from a developing country. The analysis of data gives several useful results, which can be used for future planning to reduce RTAs in developing countries. This paper also presents that how data mining model is better than other models.

Keywords: Data mining, road accident, vehicles, clusters, traffic road.

Behaviour Informatics: Capturing Value Creation in the Era of Big Data

Chi-Hung Chi

Digital Productivity Flagship, CSIRO, Australia
chihungchi@gmail.com

Abstract. Under the era of Big Data, people have been exploring ways of realizing value from data that are at their fingertips. However, it is found that while collecting data is not difficult, value creation is often a big challenge. This makes the approach of "collecting data first before knowing what to do with them" questionable. In this presentation, we discuss the current challenges of big data analytics and suggest how behaviour analytics on trajectory data can help to realize value creation from Big Data.

1 Background and Challenges

As we move to the fourth paradigm of computing – data intensive scientific discovery, numerous research efforts have been spent in building huge big data repositories. Together with data mining and machine learning research, it is hoped that better and more intelligent decisions can be made in real time.

This movement is accelerated by the advance in at least three areas. The first one is social network, where people share their views and opinions in public. The second one is cloud computing, which is an on-demand infrastructure that facilitates sharing of data, collaboration among multiple parties, and support for on-demand computational and storage infrastructure services at low cost. The third one is the internet-of-things. With the maturity of sensor technologies, trajectory movement of entities (including human and things) can now be monitored in real time at low cost. However, gaining access to big data is only the starting point. There are still open issues that need to be addressed in the value creation process when dealing with big data.

One result of the big data mega trend is the building of huge data repositories around the world. In Australia, the government has been pushing for sharing bureau data through spatial information platforms. It is true that data are collected and can be made available to users, but how to make sense out of these data practically and economically is still a mystery to be explored. Without value creation, the high maintenance cost of these repositories cannot be justified, and the motivation for data providers to update their data inside will also disappear.

In the past few years, sensors and sensing techniques have been advancing rapidly for real time data collection with good enough accuracy. Cost of deploying these technologies is also becoming low enough to make real-time data tracking of human,

animals, and even insects (e.g. honey bees) possible. However, without efficient and effective ways to integrate and transform these trajectory data and their context information into manageable knowledge, these data are actually burdens instead of potentials to their owners.

It is true that there have been numerous research efforts in data mining and machine learning. However, most of them are focused on theoretical algorithmic study, and much less emphasis is put in the incorporation of semantic domain knowledge (in particular, the semantic definition of interdependence among various data sources) into the data mining and pattern discovery processes, and in the use of the behaviour interior dimensions such as loyalty and purchase power of customers to support self service analytics.

Related to the analytics platform, internet-of-things, service and cloud computing techniques are quite mature, and lots of machine learning algorithms are also widely available in both commercial (e.g. MatLib) and open source ("Project R") packages. However, how to put them together in a single service platform and how to compose them together automatically (this is called the vertical service composition) to provide "intelligence-as-a-service" for a given domain are still open for exploration.

2 Real Time Trajectory Data and Its Challenges in Value Creation

In the era of big data, one new important data source for analytics and value creation is the real-time behaviour trajectory data streams of entities (e.g. human) as well as their context dynamics (e.g. environmental such as air quality) that are captured through internet-of-things and sensors (in particular body sensors such as those from Android wears and location position sensors). Its value creation process is both complex and challenging because these data are in general heterogeneous and inter-dependent on each other. Furthermore, the potential number of data sources, each describing one measurement view of the behaviour dynamics of an entity/event, is in theory, infinite.

Traditional data mining and machine learning approaches from computer science often try to explore co-occurrence patterns and inter-relationship among trajectory data. However, this is usually done without making full use of the interdependence defined by their implicit semantic meaning and domain knowledge. Heterogeneity of data adds another level of complication because quantification measures such as distance are not uniformly and consistently defined across different data types. On the other hand, although domain experts have full knowledge on the semantics of data, they are often not as knowledgeable as computer scientists when dealing with the real time computation on trajectory data streams. This result in the first challenge, how to use data mining / machine learning techniques and domain knowledge together to effectively define and discover the inter-relationships among different trajectory data sources and to perform effective behaviour analysis.

As trajectory-driven behaviour analytics is gaining its recognition in different business and industry sectors, the expectation of decision makers also goes beyond what traditional analytics that mainly focus on statistical summaries and association/patterns discovery of transactional/measurable behaviour exterior dimensions often provide. Ultimately, what decision makers want is the deep insight about the behaviour interior

knowledge dimensions of entities, by incorporating domain knowledge into the knowledge discovery processes. As an example, the owner of an online shop wants to know not only the "bestselling products of the week", but also the "loyalty", "purchase power", "experience", and "satisfaction" of customers. This results in the second challenge, how to quantify behaviour interior dimensions from exterior transactional (or physically measured) trajectory data and to discover their inter-relationships and relative importance for effective and efficient behaviour analysis.

3 Research Topics in Behaviour Analytics

To achieve this goal, the following is a list of sample research topics for behaviour analytics:

- Effective and efficient deployment of high resolution location tracking network (using Blue-Tooth LE, WiFi-RFIDs, UWB, and Electromagnetic Field) for entities in both indoor and outdoor environment. This forms the basis for behaviour trajectory data tracking and capturing.
- Semantic enrichment of behaviour trajectory data of entities through aggregation of raw trajectory data with their contextual data dynamics, followed by domain knowledge-driven transformation to form behaviour interior dimensions knowledge. This is the data aggregation, integration, and transformation aspects of behaviour analytics; it incorporates domain knowledge into the behaviour trajectory data to create behaviour interior dimensions knowledge as well as to define the interdependence relationship among them.
- Discovery of interdependence relationship among trajectory-driven behaviour data (exterior) and knowledge streams (interior) using data mining techniques. This addresses the interdependence relationships of trajectory data and knowledge streams from the run-time dynamics aspect.
- Coupling interdependence relationships of behaviour trajectory data and knowledge streams into data mining and pattern discovery processes for deep behaviour understanding and prediction. This gives a much better understanding on why things occur; it also gives potentials for future behaviour prediction.
- Design and implementation of a behaviour analytics service system that serves as a publishing, management and operation platform for: (i) software services, (ii) raw trajectory data services, (iii) semantically annotated behaviour trajectory data services (both individuals and collective), (iv) behaviour knowledge services (both individuals and collective), and (v) infrastructure services. Tools to facilitate composition and orchestration of all these services with QoS assurance using public cloud infrastructure such as Amazon EC2 should be developed. Also, automatic matching of behaviour trajectory data/knowledge services with machine learning/data mining algorithms based on their features should also be supported on this platform.

On the Relation of Probability, Fuzziness, Rough and Evidence Theory

Rolly Intan

Petra Christian University
Department of Informatics Engineering
Surabaya, Indonesia
rintan@petra.ac.id

Abstract. Since the appearance of the first paper on fuzzy sets proposed by Zadeh in 1965, the relationship between probability and fuzziness in the representation of uncertainty has been discussed among many people. The question is whether probability theory itself is sufficient to deal with uncertainty. In this paper the relationship between probability and fuzziness is analyzed by the process of perception to simply understand the relationship between them. It is clear that probability and fuzziness work in different areas of uncertainty. Here, fuzzy event in the presence of probability theory provides *probability of fuzzy event* in which fuzzy event could be regarded as a generalization of crisp event. Moreover, in rough set theory, a rough event is proposed representing two approximate events, namely lower approximate event and upper approximate event. Similarly, in the presence of probability theory, rough event can be extended to be *probability of rough event*. Finally, the paper shows and discusses relation among lower-upper approximate probability (probability of rough events), belief-plausibility measures (evidence theory), classical probability measures, probability of generalized fuzzy-rough events and probability of fuzzy events.

Keywords: Probability, Rough Sets, Fuzzy Sets, Evidence Theory.

Contents

Invited Paper

Fuzzy Logic and Control System

Genetic Algorithm and Heuristic Approaches

Artificial Intelligence and Machine Learning

Similarity-Based Models

Classification and Clustering Techniques

Intelligent Data Processing

Feature Extraction

Image Recognition

Visualization Techniques

Intelligent Network

Cloud and Parallel Computing

Strategic Planning

Intelligent Applications

Intelligent Systems for Enterprise, Government and Society

Invited Paper

On the Relation of Probability, Fuzziness, Rough and Evidence Theory

Rolly Intan

Petra Christian University
Department of Informatics Engineering
Surabaya, Indonesia
rintan@petra.ac.id

Abstract. Since the appearance of the first paper on fuzzy sets proposed by Zadeh in 1965, the relationship between probability and fuzziness in the representation of uncertainty has been discussed among many people. The question is whether probability theory itself is sufficient to deal with uncertainty. In this paper the relationship between probability and fuzziness is analyzed by the process of perception to simply understand the relationship between them. It is clear that probability and fuzziness work in different areas of uncertainty. Here, fuzzy event in the presence of probability theory provides *probability of fuzzy event* in which fuzzy event could be regarded as a generalization of crisp event. Moreover, in rough set theory, a rough event is proposed representing two approximate events, namely lower approximate event and upper approximate event. Similarly, in the presence of probability theory, rough event can be extended to be *probability of rough event*. Finally, the paper shows and discusses relation among lower-upper approximate probability (probability of rough events), belief-plausibility measures (evidence theory), classical probability measures, probability of generalized fuzzy-rough events and probability of fuzzy events.

Keywords: Probability, Rough Sets, Fuzzy Sets, Evidence Theory.

1 Introduction

Since the appearance of the first paper on fuzzy sets proposed by Zadeh in 1965, the relationship between probability and fuzziness in the representation of uncertainty has been discussed among many people. The main problem is whether probability theory itself is sufficient to deal with uncertainty. This issue has been widely discussed in many papers and written by Nguyen [15], Kosko [14] Zadeh [20], [21] and so on.

In this work, again just try to understand the relationship between probability and fuzziness using the process of perception by humans. In the process of perception, the subject (human, computer, robot, etc.) tries to recognize and describe a given object (anything, like human, plant, animal, event, condition, etc.).

To conduct a successful process of perception, subject requires adequate knowledge. On the other hand, object requires a clear definition. However, human (as subject) do not know what happens in the future and has also limited knowledge. In other words, humans

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 3–16, 2015.
DOI: 10.1007/978-3-662-46742-8_1

are not omniscient. In this case, the subject is in a non-deterministic situation in performing a perception. On the other hand, most objects (shape, feeling, emotion, etc.) cannot generally be clearly defined. Therefore, the perception process is in uncertain situation.

To summarize the relationship between subject and object in the process of perception, there are four possible situations as follows [10].

- If the subject has sufficient knowledge and the object has a clear definition, it becomes a certainty.
- If the subject has sufficient knowledge and object has unclear definition, it comes to the situation of fuzziness. In general, fuzziness, also called deterministic uncertainty, may occur in the situation when one is able to subjectively determine or describe a given object, but somehow the object does not have a specific or clear definition. For example, a man describes a woman as a beautiful woman. Obviously definition of a beautiful woman is unclear, uncertain and subjective. The man, however is convinced of what he describes someone as a pretty woman.
- If the subject is not in having sufficient knowledge and object has a clear definition, it comes to the situation of randomness. Randomness is usually called non-deterministic uncertainty because subject cannot determine or describe a given object clearly although the object has a clear definition. Here, probability theory was developed for dealing with the random experiment. For example, in throwing a dice, even though there are six possible defined result of outcomes, one cannot ensure outcome of dice. Another example, in solving multiple choice problem, because of his limited knowledge, a student may not be assured to choose an answer out of 4 possible answers.
- If the subject is in insufficient knowledge and object definition is unclear, it comes to be a probability of fuzzy event [20]. In this situation, both the probability and fuzziness are combined. For example, how to predict an ill-defined event: "Tomorrow will be a warm day." Speaking of morning is talk about a future in which the subject cannot determine what happens in the future. The situation must be addressed by probability. However, "hot" is a subjectively defined event (called fuzzy event). Therefore, the event is regarded to be a probability of fuzzy event.

From these four situations obviously seen that the probability and the fuzziness in work in different areas of uncertainty. Probability theory itself is not sufficient to deal with ill-defined event or fuzzy event. Instead, probability and Fuzziness should be considered as a complementary tool.

In probability, set theory is used to provide a language for modeling and describing random experiments. In (classical) theory of sets, subsets of the sample space of an experiment are referred to as crisp events. Fuzzy set theory proposed by Zadeh in 1965, is regarded as a generalization of (classical) set theory in which fuzzy sets is to represent deterministic uncertainty by a class or classes that do not have sharp defined boundary [21].

In fuzzy set theory, an ill-defined event, called fuzzy event can be described in the presence of probability theory called probability of fuzzy event [20] in which fuzzy event could be regarded as a generalization of crisp event. Conditional probability as an important property in the theory of probability usually used in generating inference rule can be extended to the conditional probability of fuzzy event. In the situation of uniform distribution probability, conditional probability of fuzzy event can be simplified to be what we call fuzzy conditional probability relations as proposed in [3] and [4] to calculate the similarity of two fuzzy labels (sets).

Similarly, rough sets theory generalizes the classical set theory by studying sets with imprecise boundaries. A rough set [16], which is characterized by a pair of approximations, can be seen as an approximate representation of a given classical set in terms of two subsets derived from a partition in the universe as proposed in [12], [13], [16] and [19]. By the theory of rough sets, rough event which is consists of lower approximate event and upper approximate event, in the presence of probability theory provides probability of rough event. Therefore, rough event could be also considered as an approximation of a given crisp event. Moreover, the probability of rough event presents a semantic formulation to what called the interval probability. Formulation of the interval probability is useful to represent the worst and the best probabilities of an event for supporting decision making process. In this paper, special attention is focused to discuss conditional probability of rough event and proved that it satisfied some properties.

In addition, a generalized fuzzy rough set as proposed in [5] and [7] is considered as an approximation of a given fuzzy set on a given fuzzy covering. Since fuzzy set and fuzzy covering generalize crisp set and crisp partition respectively, the generalized fuzzy rough set is regarded as a generalization of rough fuzzy sets and rough fuzzy sets as proposed by Dubois and Prade in [2]. Therefore, a generalized fuzzy rough event represented by generalized fuzzy rough set, in the presence of probability, provides probability of generalized fuzzy rough event. The generalized fuzzy-rough event is represented in four approximates, namely lower maximum of fuzzy event, lower minimum of fuzzy event, upper maximum of fuzzy event and upper minimum of fuzzy event.

Finally, we show and discuss relation among lower-upper approximate probability (probability of rough events), belief-plausibility measures (evidence theory), classical probability measures, probability of generalized fuzzy-rough events and probability of fuzzy events.

2 Probability of Fuzzy Event

Probability theory is based on the paradigm of a random experiment in which outcome of the experiment cannot be predicted with certainty, before running the experiment. In other words, as discussed in the previous section, in the situation of probability, subject does not have sufficient knowledge to determine outcome of the experiment. In probability, set theory is used to provide a language for modeling and describing random experiments. The sample space of a random experiment corresponds to the universal set. In (classical) theory of sets, subsets of the sample space of an experiment are used to represent crisp events.

To represent an ill-defined event, crisp event should be generalized to the fuzzy event in which the fuzzy sets used to represent fuzzy event. Formally, probability of fuzzy event is showed in the following definition [20].

Definition 1. Let (U,F,P) be regarded as a probability space in which U is the sample space, F presents sigma algebra of events and P is a probability of an event over U. Then, a fuzzy event $A \in F$ is represented by a fuzzy set A on U whose membership function given by $\mu_A : U \rightarrow [0, 1]$. The probability of fuzzy event A is defined by the following equations:

— continuous sample space:

$$P(A) = \int_U \mu_A(u) dP = \int_U \mu_A(u) \cdot p(u) du \quad (1)$$

— discrete sample space:

$$P(A) = \sum_U \mu_A(u) \cdot p(u) \quad (2)$$

where $p(u)$ is a probability distribution function of an element $u \in U$.

For example, arbitrarily given a sentence "John ate *a few* eggs for breakfast". Here, we have insufficient knowledge to know exactly how many eggs John ate for breakfast. Instead, probability distribution function of "John ate $u \in U$ egg(s) for breakfast" is arbitrarily shown by Table 1.

Table 1. Probability Distribution of u

U	1	2	3	4	5	6	• • •
$p(u)$	0.33	0.27	0.2	0.13	0.07	0	• • •

A meaningful fuzzy label, "*a few*" represented by a fuzzy event is arbitrarily given by a fuzzy set, $\mu_{afew} = \{1/1, 0.6/2, 0.2/3\}$, where $\mu_{afew}(2) = 0.6$. Probability of "John ate *a few* eggs for breakfast", denoted by P (*a few*), is calculated as follow.

$$P(a\ few) = 1 \times 0.33 + 0.6 \times 0.27 + 0.2 \times 0.2 = 0.532.$$

Some basic concepts and operations relating to the fuzzy sets are given by the following operations. Given A and B are two fuzzy sets on U [21],

- Union: $\mu_{A \cup B}(u) = \max[\mu_A(u), \mu_B(u)]$,
- Complement: $B = \neg A \Leftrightarrow \mu_B(u) = 1 - \mu_A(u), \forall u$,
- Intersection: $\mu_{A \cap B}(u) = \min[\mu_A(u), \mu_B(u)]$,
- Sum: $\mu_{A \oplus B}(u) = \mu_A(u) + \mu_B(u) - \mu_A(u) \cdot \mu_B(u)$.
- Equality: $A = B \Leftrightarrow \mu_A(u) = \mu_B(u), \forall u$,
- Containment: $A \subset B \Leftrightarrow \mu_A(u) \leq \mu_B(u), \forall u$,
- Product: $\mu_{AB}(u) = \mu_A(u) \cdot \mu_B(u)$,

It can be verified obviously that the probability of fuzzy event satisfies certain properties. Given A and B are two fuzzy sets on U,

1. $P(A \oplus B) = P(A) + P(B) - P(A \cdot B)$,
2. $A \subset B \Rightarrow P(A) \leq P(B)$,
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
4. $P(A \cap \neg A) \geq 0$.
5. $P(A \cup \neg A) \leq 1$,

1, 2 and 3 prove that probability of fuzzy event satisfies additivity axiom of sum, monotonicity and additivity axiom of union, respectively. However, it does not satisfy law of non-contradiction and law of excluded middle as clearly seen in (4) and (5).

We now turn to the notion of conditional probability of fuzzy events. Conditional probability of an event is the probability of the event given that another event has already occurred. The following equation show relationship between the conditional and unconditional probability.

$$\mathrm{P}(A|B) = \mathrm{P}(A \cap B)/\mathrm{P}(\mathrm{B}),$$

where B is an event such that $P(B) \neq 0$.

In discrete sample space, the conditional probability of fuzzy event could be defined as follows. Given A and B are two fuzzy sets on U,

$$P(A|B) = \frac{\sum_U \min\left[\mu_A(u), \mu_B(u)\right] \cdot p(u)}{\sum_U \mu_B(u) \cdot p(u)}, \forall u \in U, \tag{3}$$

where $\sum_U \mu_B(u) \cdot p(u) > 0$. It can be proved that several properties are satisfied in the conditional probability of fuzzy event. Given A and B be two fuzzy sets on U,

1. Normalization: $\mathrm{P}(A|B) + \mathrm{P}(\neg A|B) \geq 1$,
2. Total Probability: If $\{B_k \,|k \in \mathrm{N}_n\}$ are pairwise disjoint, crisp and exhaustive events,

   i.e., $\mathrm{P}(B_i \cap B_j) = 0$ for $i \neq j$ and $\cup B_k = U$, then:

$$P(A) = \sum_k P(B_k) \cdot P(A|B_k),$$

3. Bayes Theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

Furthermore, the relationship between A and B in conditional probability of fuzzy event can be represented in three conditions as follows.

- Negative correlation:

$$\mathrm{P}(\mathrm{A}|\mathrm{B}) < \mathrm{P}(\mathrm{A}) \Leftrightarrow \mathrm{P}(\mathrm{B}|\mathrm{A}) < \mathrm{P}(\mathrm{B}) \Leftrightarrow \mathrm{P}(\mathrm{A} \cap \mathrm{B}) < \mathrm{P}(\mathrm{A}) \times \mathrm{P}(\mathrm{B}),$$

- Positive correlation:

$$\mathrm{P}(\mathrm{A}|\mathrm{B}) > \mathrm{P}(\mathrm{A}) \Leftrightarrow \mathrm{P}(\mathrm{B}|\mathrm{A}) > \mathrm{P}(\mathrm{B}) \Leftrightarrow \mathrm{P}(\mathrm{A} \cap \mathrm{B}) > \mathrm{P}(\mathrm{A}) \times \mathrm{P}(\mathrm{B}),$$

- Independent correlation:

$$\mathrm{P}(\mathrm{A}|\mathrm{B}) = \mathrm{P}(\mathrm{A}) \Leftrightarrow \mathrm{P}(\mathrm{B}|\mathrm{A}) = \mathrm{P}(\mathrm{B}) \Leftrightarrow \mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}) \times \mathrm{P}(\mathrm{B}).$$

In the situation of uniform distribution, the probability distribution function $p(u) = 1 / |U|$, is considered as a constant variable. Therefore, the conditional probability of A given B is more simply defined by eliminating $p(u)$ as given by:

$$P(A|B) = \frac{\sum_U \min\left[\mu_A(u), \mu_B(u)\right]}{\sum_U \mu_B(u)}, \forall u \in U, \tag{4}$$

In [3] and [4], the Equation (4) called *fuzzy conditional probability relation* is used to calculate degree of similarity relationship between two fuzzy labels (sets).

3 Probability of Rough Event

Rough set is considered as a generalization of crisp set by studying sets with imprecise boundaries. A rough set, characterized by a pair approximations, called lower approximation and upper approximation, can be seen as an approximate representation of a given crisp set in terms of two subsets derived from a partition in the universe as explained in [12], [13], [16] and [19].

The concept of rough sets can be defined precisely as follows. Let U denotes a finite and non-empty universe, and let R be an equivalence relation on U. The equivalence relation R induces a partition of the universe. The partition is also referred to as the quotient set and is denoted by U/R. Suppose $[u]_R$ is the equivalence class in U/R that contains $u \in U$. A rough set approximation of a subset $A \subseteq U$ is a pair of lower and upper approximations.

Formally, rough sets may be defined precisely as follows. Let U denotes a non-empty and finite universe. R is an equivalence relation on U. The equivalence relation R induces a partition of the universe. The partition is also known as the quotient set, and it is denoted by U/R. Given $[u]_R$ is the equivalence class in U/R that consists of u ∈ U. A rough set of a subset $A \subseteq U$ is represented by a pair of lower and upper approximations as given by the following equations. The lower approximation,

$$Lo(A) = \{u \in U | [u]_R \subseteq A\} = \bigcup\{[u]_R \in U/R | [u]_R \subseteq A\},$$

is the union of all equivalence classes in U/R that are subsets of A. The upper approximation,

$$Up(A) = \{u \in U | [u]_R \cap A \neq \emptyset\} = \bigcup\{[u]_R \in U/R | [u]_R \cap A \neq \emptyset\},$$

is the union of all equivalence classes in U/R that overlap with A. Similarly, by rough set, an event can be described into two approximate rough events, namely lower approximate event and upper approximate event. Rough event can be considered as the generalization and approximation of a given crisp event. Probability of rough event is then defined by the following equations.

Definition 2. Let (U,F,P) be regarded as a probability space in which U is the sample space. F represents sigma algebra of events, and P is a probability measure over U. Then, a rough event of $A = [Lo(A), Up(A)] \in F^2$ is given by a pair of approximations, called lower approximation and upper approximation of $A \subseteq U$. The probability of rough event A is defined by an interval probability $[P(Lo(A)), P(Up(A))]$, where $P(Lo(A))$ and $P(Up(A))$ are lower probability and upper probability, respectively.

– Lower probability:

$$P(Lo(A)) = \sum_{\{u \in U | [u]_R \subseteq A\}} p(u) = \sum_{\cup\{[u]_R \in U/R | [u]_R \subseteq A\}} P([u]_R), \quad (5)$$

– Upper probability:

$$P(Up(A)) = \sum_{\{u \in U | [u]_R \cap A \neq \emptyset\}} p(u) = \sum_{\cup\{[u]_R \in U/R | [u]_R \cap A \neq \emptyset\}} P([u]_R), \quad (6)$$

where $p(u)$ is a probability distribution function of element $u \in U$.

Lower and upper probabilities in Definition 2 can be regarded as an interval probability. By combining with other set-theoretic operators such as ¬, ∪ and ∩, we have the following results of properties:

1. $P(Lo(A)) \leq P(A) \leq P(Up(A))$,
2. $A \subseteq B \Leftrightarrow [P(Lo(A)) \leq P(Lo(B)), P(Up(A) \leq P(Up(B))]$,
3. $\mathrm{P}(Lo(\neg A)) = 1 - \mathrm{P}(Lo(A)), \mathrm{P}(Up(\neg A)) = 1 - \mathrm{P}(Up(A))$,
4. $\mathrm{P}(\neg Lo(A)) = \mathrm{P}(Up(\neg A)), \mathrm{P}(\neg Up(A)) = \mathrm{P}(Lo(\neg A))$,
5. $\mathrm{P}(Lo(U)) = \mathrm{P}(U) = \mathrm{P}(Up(U)) = 1, \mathrm{P}(Lo(\emptyset)) = \mathrm{P}(\emptyset) = \mathrm{P}(Up(\emptyset)) = 0$,
6. $\mathrm{P}(Lo(A \cap B)) = \mathrm{P}(Lo(A) \cap Lo(B)), \mathrm{P}(Up(A \cap B)) \leq \mathrm{P}(Up(A) \cap Up(B))$,
7. $\mathrm{P}(Lo(A \cup B)) \geq \mathrm{P}(Lo(A)) + \mathrm{P}(Lo(B)) - \mathrm{P}(Lo(A \cap B))$,
8. $\mathrm{P}(Up(A \cup B)) \leq \mathrm{P}(Up(A)) + \mathrm{P}(Up(B)) - \mathrm{P}(Up(A \cap B))$,
9. $\mathrm{P}(A) \leq \mathrm{P}(Lo(Up(A))), \mathrm{P}(A) \geq \mathrm{P}(Up(Lo(A)))$,
10. $\mathrm{P}(Lo(A)) = \mathrm{P}(Lo(Lo(A))), \mathrm{P}(Up(A)) = \mathrm{P}(Up(Up(A)))$,
11. $\mathrm{P}(Lo(A) \cup Lo(\neg A)) \leq 1, \mathrm{P}(Up(A) \cup Up(\neg A)) \geq 1$,
12. $\mathrm{P}(Lo(A) \cap Lo(\neg A)) = 0, \mathrm{P}(Up(A) \cap Up(\neg A)) \geq 0$.

Conditional probability of rough event could be considered in the following four combinations of formulations: Given A, B ⊆ U, conditional probability of A given B is given by

$$P(Lo(A)|Lo(B)) = \frac{P(Lo(A) \cap Lo(B))}{P(Lo(B))}, \tag{7}$$

$$P(Lo(A)|Up(B)) = \frac{P(Lo(A) \cap Up(B))}{P(Up(B))}, \tag{8}$$

$$P(Up(A)|Lo(B)) = \frac{P(Up(A) \cap Lo(B))}{P(Lo(B))}, \tag{9}$$

$$P(Up(A)|Up(B)) = \frac{P(Up(A) \cap Up(B))}{P(Up(B))}. \tag{10}$$

It can be proved that the equations are also satisfied some relations as given by:

$$P(Lo(A) \cap Lo(B)) \leq P(Up(A) \cap Lo(B)) \Rightarrow P(Lo(A)|Lo(B)) \leq P(Up(A)|Lo(B))$$

$$P(Lo(A) \cap Up(B)) \leq P(Up(A) \cap Up(B)) \Rightarrow P(Lo(A)|Up(B)) \leq P(Up(A)|Up(B))$$

Similarly, it can also be verified that they satisfy some properties:

1. Normalization:

   — $\mathrm{P}(\mathrm{Lo(A)}|\mathrm{Lo(B)}) + \mathrm{P}(\mathrm{Lo}(\neg \mathrm{A})|\mathrm{Lo(B)}) \leq 1$,
   — $P(Lo(A)|Up(B)) + P(Lo(\neg A)|Up(B)) \leq 1$,
   — $P(Up(A)|Lo(B)) + P(Up(\neg A)|Lo(B)) \geq 1$,
   — $P(Up(A)|Up(B)) + P(Up(\neg A)|Up(B)) \geq 1$.

2. Total Probability If $\{B_k | k \in \mathrm{N}_n\}$ are pairwise disjoint, crisp and exhaustive events, i.e., $P(B_i \cap B_j) = 0$ for $i \neq j$ and $\cup B_k = U$, then:

- $P(Lo(A)) \geq \sum_k P\big(Lo(B_k)\big) \cdot P\big(Lo(A)\big|Lo(B_k)\big)$,
- $P(Lo(A)) \leq \sum_k P\big(Up(B_k)\big) \cdot P\big(Lo(A)\big|Up(B_k)\big)$,
- $P(Up(A)) \geq \sum_k P\big(Lo(B_k)\big) \cdot P\big(Up(A)\big|Lo(B_k)\big)$,
- $P\big(Up(A)\big) \leq \sum_k P\big(Up(B_k)\big) \cdot P\big(Up(A)\big|Up(B_k)\big)$.

3. Bayes Theorem:

$$P\big(Lo(A)\big|Lo(B)\big) = \frac{P(Lo(B)|Lo(B)) \cdot P(Lo(A))}{P(Lo(B))},$$

$$P\big(Lo(A)\big|Up(B)\big) = \frac{P(Up(B)|Lo(A)) \cdot P(Lo(A))}{P(Up(B))},$$

$$P\big(Up(A)\big|Lo(B)\big) = \frac{P(Lo(B)|Up(A)) \cdot P(Up(A))}{P(Lo(B))},$$

$$P\big(Up(A)\big|Up(B)\big) = \frac{P(Up(B)|Up(A)) \cdot P(Up(A))}{P(Up(B))}.$$

We may also consider other definitions of conditional probability of rough event as given by the following equations. Given $A, B \subseteq U$, conditional probability of A given B can be also defined by

$$P_1(A|B) = \frac{P(Lo(A \cap B))}{P(Lo(B))}, \tag{11}$$

$$P_2(A|B) = \frac{P(Lo(A \cap B))}{P(Up(B))}, \tag{12}$$

$$P_3(A|B) = \frac{P(Up(A \cap B))}{P(Lo(B))}, \tag{13}$$

$$P_4(A|B) = \frac{P(Up(A \cap B))}{P(Up(B))}. \tag{14}$$

Also, it can be proved that the above equations satisfy some relations as follows.

- $P_2(A|B) \leq P_1(A|B) \leq P_3(A|B)$,
- $P_4(A|B) \leq P_3(A|B)$,
- $P_2(A|B) \leq P_4(A|B)$,
- $P\,(Lo(A \cap B)) = P\,(Lo(A) \cap Lo(B)) \Rightarrow P_1(A|B) = P\,(Lo(A)|Lo(B))$.

Moreover, they also satisfy some properties of conditional probability:

1. Normalization:

- $P_1(A|B) + P_1(\neg A|B) \leq 1$,
- $P_2(A|B) + P_2(\neg A|B) \leq 1$,
- $P_3(A|B) + P_3(\neg A|B) \geq 1$,
- $P_4(A|B) + P_4(\neg A|B) \geq 1$.

2. Total Probability If $\{B_k \,|k \in N_n\}$ are pairwise disjoint, crisp and exhaustive events,

i.e., $P\,(B_i \cap B_j) = 0$ for $i \neq j$ and $\cup B_k = U$, then:

- $P(Lo(A)) \geq \sum_k P\big(Lo(B_k)\big) \cdot P_1(A|B_k)$,
- $P(Lo(A)) \geq \sum_k P\big(Lo(B_k)\big) \cdot P_2(A|B_k)$,
- $P(Lo(A)) \leq \sum_k P\big(Lo(B_k)\big) \cdot P_3(A|B_k)$,
- $P\big(Lo(A)\big) \leq \sum_k P\big(Lo(B_k)\big) \cdot P_4(A|B_k)$.

3. Bayes Theorem:

$$P_1(A|B) = \frac{P_1(B|A) \cdot P(Lo(A))}{P(Lo(B))},$$

$$P_2(A|B) = \frac{P_2(B|A) \cdot P(Lo(A))}{P(Up(B))},$$

$$P_3(A|B) = \frac{P_3(B|A) \cdot P(Up(A))}{P(Lo(B))},$$

$$P_4(A|B) = \frac{P_4(B|A) \cdot P(Up(A))}{P(Up(B))}.$$

4 Probability of Generalized Fuzzy-Rough Event

A generalized fuzzy rough set is an approximation of a given fuzzy set on a given fuzzy covering. Since fuzzy set generalizes crisp set and covering generalizes partition, fuzzy covering is regarded as the most generalized approximation space. Fuzzy covering might be considered as a case of *fuzzy granularity* in which similarity classes as a basis of constructing the covering are regarded as fuzzy sets. Alternatively, a fuzzy covering might be constructed and defined as follows [6].

A generalized fuzzy rough set is an approximation of a given fuzzy set in a given fuzzy covering. Since fuzzy set and covering generalized crisp set and partition, respectively, fuzzy covering is considered the most generalized approximation space. Fuzzy covering can be considered as a case of fuzzy granularity. In this case, similarity classes as used in constructing the covering are considered as fuzzy sets. The following definition shows an alternative definition in constructing a fuzzy covering [6].

Definition 3. Let $U = \{u_1, ..., u_n\}$ be regarded as a universe. A fuzzy covering of U is defined by a family of fuzzy subsets or fuzzy classes of C, denoted by $C = \{C_1, C_2, ..., C_m\}$, and it satisfies

$$\sum_{i=1}^{m} \mu_{C_i}(u_k) \geq 1, \forall k \in N_n \tag{15}$$

$$0 < \sum_{k=i}^{n} \mu_{C_i}(u_k) < n, \forall i \in N_m \tag{16}$$

where $\mu_{Ci}(u_k) \in [0, 1]$ and m is a positive integer.

Let A be a fuzzy set on fuzzy covering as defined in Definition 3. A generalized fuzzy rough set A is then defined by the following definition.

Definition 4. Let U be regarded as a non-empty universe. $C = \{C_1, C_2, ..., C_m\}$ is a fuzzy covering on U. Given A be a fuzzy set on U, $Lo(A)_M$, $Lo(A)_m$, $Up(A)_M$ and $Up(A)_m$ are defined as minimum lower approximate, maximum lower approximate, minimum upper approximate and maximum upper approximate fuzzy set of A, respectively, as follows.

$$\mu_{Lo(A)_m}(y) = \inf_{\{i|\mu_{C_i}(y)>0\}} \inf_{\{z\in U|\mu_{C_i}(z)>0\}}\{\Psi(i,z)\}, \quad (17)$$

$$\mu_{Lo(A)_M}(y) = \sup_{\{i|\mu_{C_i}(y)>0\}} \inf_{\{z\in U|\mu_{C_i}(z)>0\}}\{\Psi(i,z)\}, \quad (18)$$

$$\mu_{Up(A)_m}(y) = \inf_{\{i|\mu_{C_i}(y)>0\}} \sup_{\{z\in U\}}\{\Psi(i,z)\}, \quad (19)$$

$$\mu_{Up(A)_M}(y) = \sup_{\{i|\mu_{C_i}(y)>0\}} \sup_{\{z\in U\}}\{\Psi(i,z)\}, \quad (20)$$

where $\Psi(i,z) = \min(\mu_{C_i}(z), \mu_A(z))$, for short.

Therefore, a given fuzzy set *A* is approximated into four approximate fuzzy sets. It can be proved that relationship among these approximations satisfy a partial order as follows.

$$Lo(A)_m \subseteq Lo(A)_M \subseteq Up(A)_M, Lo(A)_m \subseteq Up(A)_m \subseteq Up(A)_M, Lo(A)_M \subseteq A$$

The property of iterative is also applied for almost all approximate fuzzy sets except for $Lo(A)_M$ as shown in the following relations.

- $Lo(A)_{m*} \subseteq \cdots \subseteq Lo(Lo(A)_m)_m \subseteq Lo(A)_m$,
- $Up(A)_m \subseteq Up(Up(A)_m)_m \subseteq \cdots \subseteq Up(A)_{m*}$,
- $Up(A)_M \subseteq Up(Up(A)_M)_M \subseteq \cdots \subseteq Up(A)_{M*}$,

where $Lo(A)_{m*}$, $Up(A)_{m*}$ and $Up(A)_{M*}$ are regarded as the lowest approximation of $Lo(A)_m$, the uppermost approximation of $Up(A)_m$ and the uppermost approximation of $Up(A)_M$, respectively. Thus, in our concept of the generalized fuzzy rough set, a given fuzzy event can be represented into four fuzzy events, called *generalized fuzzy-rough event*. In the relation to probability theory, probability of generalized fuzzy-rough event is then defined in Definition 5 as follows.

Definition 5. Let (*U, F, P*) be regarded as a probability space in which *U* is the sample space, *F* is defined as sigma algebra of events, and *P* is a probability measure over *U*. Then, a generalized fuzzy-rough event of $\mathbf{A} = [Lo(A)_m, Lo(A)_M, Up(A)_m, Up(A)_M] \in F^4$ are considered as fuzzy approximate events of *A*, where *A* is a given fuzzy set on *U*. The probability of generalized fuzzy-rough event *A* is then defined by a quadruplet $[P(Lo(A)_m), P(Lo(A)_M), P(Up(A)_m), P(Up(A)_M)]$ as given by the following equations.

$$P(Lo(A)_m) = \sum_U \mu_{Lo(A)_m}(u) \cdot p(u), \quad (21)$$

$$P(Lo(A)_M) = \sum_U \mu_{Lo(A)_M}(u) \cdot p(u), \quad (22)$$

$$P(Up(A)_m) = \sum_U \mu_{Up(A)_m}(u) \cdot p(u), \quad (23)$$

$$P(Up(A)_M) = \sum_U \mu_{Up(A)_M}(u) \cdot p(u), \quad (24)$$

where $p(u)$ is defined as a probability distribution function of element $u \in U$.

By combining with other set-theoretic operators such as ¬, ∪ and ∩, we have the following results of properties:

- $P(Lo(A)_m) \leq P(Lo(A)_M) \leq P(Up(A)_M)$,
- $P(Lo(A)_M) \leq P(A)$,
- $P(Lo(A)_m) \leq P(Up(A)_m) \leq P(Up(A)_M)$,
- $A \subseteq B \Rightarrow [P(Lo(A)_m) \leq P(Lo(B)_m),\ P(Lo(A)_M) \leq P(Lo(B)_M),\ P(Up(A)_m) \leq P(Up(B)_m),\ P(Up(A)_M) \leq P(Up(B)_M)]$,
- $P(Lo(U)_\lambda) \leq 1,\ P(Up(U)_\lambda) \leq 1$,
- $P(Lo(\emptyset)_\lambda) = P(Up(\emptyset)_\lambda) = 0$,
- $P(Lo(A \cap B)_\lambda) \leq P(Lo(A)_\lambda \cap Lo(B)_\lambda)$,
- $P(Up(A \cap B)_\lambda) \leq P(Up(A)_\lambda \cap Up(B)_\lambda)$,
- $P(Lo(A \cup B)_\lambda) \geq P(Lo(A)_\lambda) + P(Lo(B)_\lambda)) - P(Lo(A \cap B)_\lambda)$,
- $P(Up(A \cup B)_\lambda) \leq P(Up(A)_\lambda) + P(Up(B)_\lambda)) - P(Up(A \cap B)_\lambda)$,
- $P(Lo(A)_{m*}) \leq \cdots \leq P(Lo(Lo(A)_m)_m) \leq P(Lo(A)_m)$,
- $P(Lo(A)_M) = P(Lo(Lo(A)_M)_M)$,
- $P(Up(A)_m) \leq P(Up(Up(A)_m)_m) \leq \cdots \leq P(Up(A)_{m*})$,
- $P(Up(A)_M) \leq P(Up(Up(A)_M)_M) \leq \cdots \leq P(Up(A)_{M*})$,
- $P(Lo(A)_\lambda \cup Lo(\neg A)_\lambda) \leq 1$,
- $P(Lo(A)_\lambda \cap Lo(\neg A)_\lambda) \geq 0$,
- $P(Up(A)_\lambda \cap Up(\neg A)_\lambda) \geq 0$,

where $\lambda \in \{m, M\}$, for short.

5 Belief and Plausibility Measures

In evidence theory, belief and plausibility measures originally introduced by Glenn Shafer in 1976 [17] are mutually dual functions in evidence theory. This concept was strongly motivated and related to lower probability and upper probability proposed by Dempster in 1967 [1] in which all measures are subsumed in the concept of fuzzy measure proposed by Sugeno in 1977 [18]. Belief and plausibility Measures can be represented by a single function, called basic probability assignment, providing evidence grades for specific subsets of the universal set. In a special case when the subsets of the universal set are disjoint and each subset represents elementary set of indiscernible space, we can consider belief and plausibility measures as lower approximate probability and upper approximate probability in terms of probability of rough events as proposed in [8] and [9]. Here, lower approximate probability and upper approximate probability are considered as a special case of belief and plausibility measures, respectively in which probability of elementary set is considered as a special case of basic probability assignment. In other words, belief and plausibility measures are based on crisp-granularity in terms of a covering. On the other hand, lower approximate probability and upper approximate probability are defined on crisp-granularity in terms of disjoint partition. Here, when every elementary set has only one element of set, every probability of elementary set will be equal to probability of an element called probability distribution function as usually used in representing probability measures. Therefore, it can be verified that lower approximate probability and upper approximate probability of a given rough event will be clearly reduced into a single value of probability. Lower approximate probability and upper approximate probability as well as belief and plausibility measures are also regarded as generalization of probability measures in the

presence of crisp granularity of sample space. We may consider another generalization when the membership degree of every element of sample space in representing an event is regarded from 0 to 1. The generalization provides a concept called probability measures of fuzzy events as proposed by Zadeh in 1968 [20]. Moreover, it may then propose a more generalized probability measures by given a fuzzy event in the presence of fuzzy-granularity of sample space. The generalization called probability measures of generalized fuzzy-rough events as already discussed in previous section [5] and [7]. Belief and plausibility measures can be also represented by a single function called basic probability assignment as defined by the following [11]:

Definition 6. Given U be regarded as a universal sample space and $\mathcal{P}(U)$ be a power set of U,

$$m : \mathcal{P}(\underline{U}) \rightarrow [0, 1] \tag{25}$$

such that $\sum_{E\in\mathcal{P}(U)} m(E) = 1$ and $m(\emptyset) = 0$, where $m(E)$ expresses the degree of evidence supporting the claim that a specific element of U belongs to the set E but not to any special subset of E. There are three properties considering the definition of basic probability assignment.

1. It is not necessary that $m(U) = 1$.
2. It is not necessary that $E_1 \subset E_2 \Rightarrow m(E_1) \leq m(E_2)$.
3. There is no relationship between $m(E)$ and $m(\neg E)$.

Every $E \in \mathcal{P}(U)$ is called a *focal element* iff $m(E) > 0$. It is possible that focal elements may take overlap one to each other. Belief and Plausibility measures are then given by Equation (26) and (27), respectively. For $A \in \mathcal{P}(U)$,

$$Bel(A) = \sum_{E\subseteq A} m(E), \tag{26}$$

$$Pl(A) = \sum_{E\cap A\neq\emptyset} m(E). \tag{27}$$

It can be also verified that for all $A \in \mathcal{P}(U)$, $Bel(A) \leq Pl(A)$. As mention before, it can be proved that belief and plausibility measures are mutually dual functions as shown in the following equations.

$$Pl(A) = 1 - Bel(\neg A).$$

Similarly,

$$Bel(A) = 1 - Pl(\neg A).$$

Belief and plausibility measures are defined on a covering. Therefore, some properties are not satisfied, especially iterative properties of lower approximate probability and upper approximate probability such as given by $\mathrm{P}(A) \leq \mathrm{P}(Lo(Up(A)))$, $\mathrm{P}(A) \geq \mathrm{P}(Up(Lo(A)))$ and $\mathrm{P}(Lo(A)) = \mathrm{P}(Lo(Lo(A)))$, $\mathrm{P}(Up(A)) = \mathrm{P}(Up(Up(A)))$ as also mentioned in Section 3. Let consider,

$$Pl^{-1}(A) = \bigcup_{E\in\mathcal{P}(U),E\cap A\neq\emptyset} E \quad \text{and} \quad Bel^{-1}(A) = \bigcup_{E\in\mathcal{P}(U),E\subseteq A} E,$$

where $Pl(A)$ and $Pl^{-1}(A)$ correspond to $P(Up(A))$ and $Up(A)$, respectively. Similarly, $Bel(A)$ and $Bel^{-1}(A)$ correspond to $P(Lo(A))$ and $Lo(A)$, respectively. Hence, property of $P(Up(A)) = P(Up(Up(A)))$ can be represented as $Pl(A) = Pl(Pl^{-1}(A))$ by using expression

of plausibility measures. It can be easily verified that the property is not satisfied instead $Pl(A) \leq Pl(Pl^{-1}(A))$. Also, $P(A) \geq Pl(Bel^{-1}(A))$ in the relation to $P(A) \geq P(Up(Lo(A)))$, cannot be verified. When every elementary set has only one element, the probability of elementary set is equal to probability of the element represented by a function called probability distribution function, $p : U \rightarrow [0,1]$, which is defined on set U as usually used in probability measures. Here, lower approximate probability and upper approximate probability fuse into a single value of probability in which probability satisfies additivity axiom as an intersection area between supperadditive property of lower approximate probability and subadditive property of upper approximate probability as respectively given by the equations, $\mathrm{P}(Lo(A \cup B)) \geq \mathrm{P}(Lo(A)) + \mathrm{P}(Lo(B)) - \mathrm{P}(Lo(A \cap B))$ and $\mathrm{P}(Up(A \cup B)) \leq \mathrm{P}(Up(A)) + \mathrm{P}(Up(B)) - \mathrm{P}(Up(A \cap B))$ as also already mentioned in Section 3.

6 Conclusion

The relationship between probability and fuzziness was discussed clearly based on the process of perception. Probability and fuzziness work in different areas of uncertainty; therefore, probability theory itself is not sufficient to deal with the uncertainty in real world application. Instead, fuzziness and probability must be regarded as a complementary concepts to represent various type of uncertainty. For instance, relation between fuzziness and probability may provide a concept, called probability of fuzzy event in which fuzzy event was represented by a given fuzzy set. Here, fuzzy event and fuzzy set are considered as a generalization of crisp event and crisp set, respectively. Similarly, rough set, as another generalization of crisp set, is used to represent rough event. In the presence of probability theory, probability of rough event was also proposed as another generalization of probability measure. Conditional probability of fuzzy event and conditional probability of rough event were examined together with their some properties. A more generalized fuzzy rough set is then proposed as an approximation of a given fuzzy set in a given fuzzy covering.

Therefore, using the concept of generalized fuzzy rough set, a generalized fuzzy-rough event was proposed as the most generalization of fuzzy event as well as rough event. Probability of the generalized fuzzy-rough event was introduced along with its properties. Figure 1 shows summary of relationship among all the concepts.

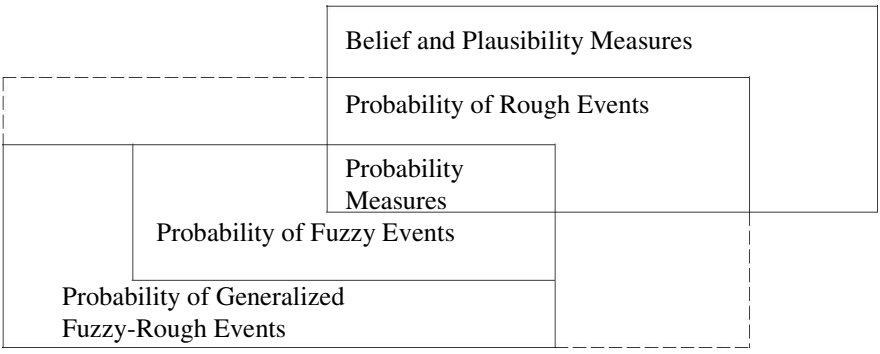


Fig. 1. Generalization based on Crisp-Granularity and Membership Function

References

1. Dempster, A.P.: Upper and Lower Probability Induced by Multivalued Mappings. Annals of Mathematical Statistics 38, 325–339 (1967)
2. Dubois, D., Prade, H.: Rough Fuzzy Sets and Fuzzy Rough Set. International Journal of General Systems 17, 191–209 (1990)
3. Intan, R., Mukaidono, M.: Application of Conditional Probability in Constructing Fuzzy Functional Dependency (FFD). In: Proceedings of AFSS 2000, pp. 271–276 (2000a)
4. Intan, R., Mukaidono, M.: Fuzzy Functional Dependency and Its Application to Approximate Querying. In: Proceedings of IDEAS 2000, pp. 47–54 (2000b)
5. Intan, R., Mukaidono, M.: Generalization of Fuzzy Rough Sets by Fuzzy Covering Based on Weak Fuzzy Similarity Relation. In: Proceeding of FSKD 2002, pp. 344–348 (2002a)
6. Intan, R., Mukaidono, M.: Degree of Similarity in Fuzzy Partition. In: Pal, N.R., Sugeno, M. (eds.) AFSS 2002. LNCS (LNAI), vol. 2275, pp. 20–26. Springer, Heidelberg (2002b)
7. Intan, R., Mukaidono, M.: Generalized Fuzzy Rough Sets By Conditional Probability Relations. International Journal of Pattern Recognition and Artificial Intelligence 16(7), 865–881 (2002c)
8. Intan, R., Mukaidono, M.: Probability of Fuzzy Event to Probability of Rough Event. In: Proceedings of FSKD 2002, pp. 549–553 (2002d)
9. Intan, R., Mukaidono, M.: A Proposal of Probability of Rough Event based on Probability of Fuzzy Event. In: Alpigini, J.J., Peters, J.F., Skowron, A., Zhong, N. (eds.) RSCTC 2002. LNCS (LNAI), vol. 2475, pp. 357–364. Springer, Heidelberg (2002e)
10. Intan, R., Mukaidono, M.: Hybrid Probabilistic Models of Fuzzy and Rough Events. Journal of Advanced Computational Intelligent 7(3), 323–329 (2003)
11. Inuiguchi, M., Tanino, T.: On Rough Sets under Generalized Equivalence Relations. In: Proceedings of International Workshop on Rough Sets and Granular Computing, pp. 167–171 (2001)
12. Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, New Jersey (1995)
13. Komorowski, J., Pawlak, Z., Polkowski, L., Skowron, A.: Rough Sets: A Tutorial. In: Pal, S.K., Skowron, A. (eds.) Rough Fuzzy Hybridization, pp. 3–98. Springer (1999)
14. Kosko, B.: Fuzziness VS. Probability. International Journal of General Systems 17, 211–240 (1990)
15. Nguyen, H.T.: On Fuzziness and Linguistic Probabilities. Journal of Mathematical Analysis and Applications 61, 658–671 (1977)
16. Pawlak, Z.: Rough sets. International Journal Computation and Information Science 11, 341–356 (1982)
17. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
18. Sugeno, M.: Fuzzy Measures and Fuzzy Integrals, a survey. In: Gupta, Saridis and Gaines, pp. 89–102 (1977)
19. Yao, Y.Y.: Two views of the theory of rough sets in finite universe. International Journal of Approximate Reasoning 15, 291–317 (1996)
20. Zadeh, L.A.: Probability Measures of Fuzzy Events. Journal of Mathematical Analysis and Applications 23, 421–427 (1968)
21. Zadeh, L.A.: Fuzzy Sets and Systems. International Journal of General Systems 17, 129–138 (1990)
22. Zadeh, L.A.: Discussion: Probability Theory and Fuzzy Logic Are Complementary Rather Than Competitive. Technometrics 37(3), 271–276 (1995)

Fuzzy Logic and Control System

A Study of Laundry Tidiness: Laundry State Determination Using Video and 3D Sensors

Hirose Daiki*, Miyoshi Tsutomu, and Maiya Kazuki

Ryukoku University, Graduate School of Science and Technology, Shiga, Japan
mijosxi@rins.ryukoku.ac.jp

Abstract. In recent years, housework automation has become popular with the rise of robot technology. However, tidying laundry is still manual. Automation by machine is difficult, because clothing is an irregular complex shape. We thought it can handle the laundry by combining depth information and color image. The purpose of this study is to develop a laundry state determination system. This study define state of laundry by dividing into 4 states, and develop laundry state determination system using RGB image and 3D information. The results of experiment of the proposed method suggest that the system was possible to accurately determination the state of the laundry by using depth information and RGB camera image with Kinect.

Keywords: Laundry Tidiness, State Determination, Kinect.

1 Introduction

In recent years, housework automation has become popular such as cleaning robot or washing and drying machine, etc. Housework automation reduce user's effort of housework, and provide a free time. Housework definition in long-term care services is classified into cooking, washing, cleaning, shopping, throw out trash, clothing organize and clothing repair. There are some agency or automation service for cooking, washing, cleaning, and shopping. For example, for cooking, microwave oven is common consumer electronics. For washing, washing and drying machine is common too. For cleaning, room cleaning robot like Roomba[1] has become popular. For shopping, shopping at online and get it by courier have become easier. However, tidying laundry which is classified as clothing organize is still manual. Automation by machine is difficult[2-4,6-10], because clothing is an irregular complex shape.

2 Automatic Laundry Tidying System

We aims to develop a laundry tidying system of closet type fitted with Kinect (Fig.1). We assume the system as following.

- The system recognized the state of laundry, and gives commands to small robots.
- Small robots can only 3 operations of grabbing, detaching and moving.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 19–26, 2015.
DOI: 10.1007/978-3-662-46742-8_2

- Large clothes such as shirts and pants tidy up by hanging without folding.
- Socks perform a pairing and put away to a particular shelf.

This study is intended to build a system for determining the state of the laundry among the laundry automatic tidying system using depth information and camera image of Kinect[5].

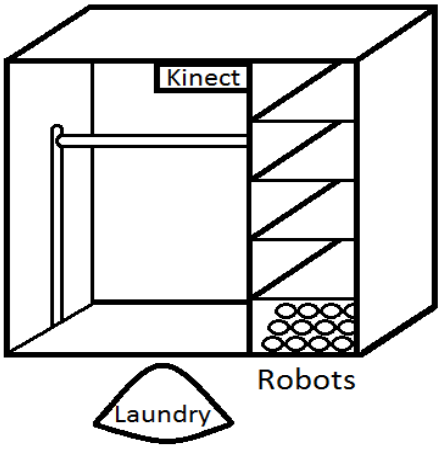


Fig. 1. Closet fitted with Kinect

3 Definition of Laundry States and Action

We defined the state of laundry in four states as follows. The first state is initial state in which the laundry is placed. The second state is recognized that a laundry was taken out one piece from the laundry pile. The third state is tidying socks which are completed pairing. The fourth state is last state that tidying of laundry is completed. This state is the same as human has finished tidying the laundry, the laundry does not exist on the floor (Fig.2).

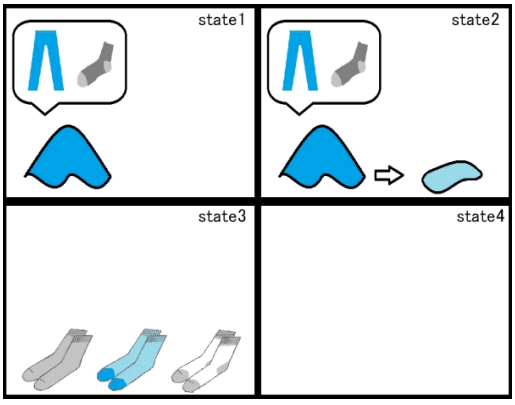


Fig. 2. State of the laundry which is divided into four

- state 1: Laundry of all types are mixed, and it is a pile.
- state 2: Laundry is taken out of one piece from laundry pile.
- state 3: Large clothes such as pants are put away. Classification of small clothes which need pairing like socks pairing are complete.
- state 4: Laundry of all has been tidied.

We defined IF-THEN rules of their behavior (Table1). State transition Diagram is as shown in Fig.3. Table1 shows transition of laundry state. "State" is the number of state that laundry state is divided into four parts. "Target" is target which is used determination of laundry state by the system. "IF" is the condition of the "Target". "THEN" is the action when "Target" conforms to the conditions of "IF". "Transition" is transition of laundry state after the action, and transition number of Fig.3.For example, if current state is 1 and the pile of laundry exist, we move a laundry. then current state will transition to state2. In this case, the state transition number is 1-1 in Fig.3. In Fig.3, the starting state is a non-recognition state.

Table 1. State Transition

state	Target	IF	THEN	Transition	
1	pile of laundry	exist	object moving process	state2	1-1
	pile of laundry	none	-	state3	1-2
2	recognized object	pants	pants tidying process	state1	2-1
	recognized object	socks	pairing process	state1	2-2
	recognized object	none	-	state1	2-3
3	small clothes	exist	socks tidying process	state3	3-1
	small clothes	none	-	state4	3-2
4	pile of laundry	exist	-	state1	4-1
	recognized object	exist	-	state2	4-2
	small clothes	exist	-	state3	4-3
	all objects	none	-	finished	

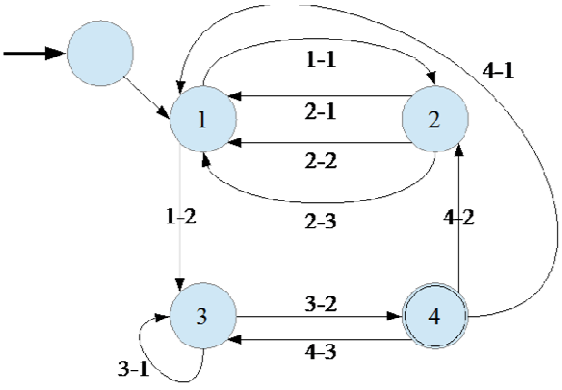


Fig. 3. State Transition Diagram

4 Target Recognition and Determination

The work flow of total process is represented as following 5 sub process.

1. Kinect set the position of 1.5m height downwardly perpendicular to the floor, and initialize Kinect.
2. Laundry pile put on the floor.
3. Recognizes the state of laundry.
4. Acting and Changing the state of laundry.
5. Repeat 3 and 4.

In this paper, we focused on the recognition of IF part. Kinect set downwardly perpendicular to the floor, and a laundry pile put on the camera screen. In addition, Do not put the object of other than laundry on the floor, and to ensure sufficient space for moving laundry. We performed the move of laundry by human operator, not by small robots.

The third process of the above is the most interested in this paper. The ditailes of this process gives as follows.

・Recognizes the state of laundry

3-1. Determination of the recognition field

3-2. If the recognized object is present, apply object recognition

3-3. If threshold less than area of the object is determined to socks

3-4. If threshold above area of the object is determined to pants

3-5. Determination of the laundry pile field

3-6. Determination of the small clothes field

3-7. Decision of the state

The first process is to set Kinect, and initialize using background subtraction for recognize the object. The process performe the following processing.

・Initialize Kinect

1-1. Storage of the background image

1-2. Get camera image and a depth of information

1-3. Detection of floor area

1-4. Create of small clothes field to the third upper side of the floor area

1-5. Create of laundry pile field by the rest of third left

1-6. Create of laundry pile field by all remaining

1-7. Extraction of laundry by the background difference method

1-8. Cut out laundry pile field, recognition field, and the small clothes field

1-9. Formed a pile by gathering the laundry in the laundry pile field

5 Experimental

We tested by implementation experimental whether or not Recognition of the IF part is performed correctly. This experiment was performed extracts the information of the laundry using the depth information and the camera image by Kinect. In this experiment use two types of pants and socks as laundry. In addition, threshold of object size is 4000pixel, because we determined to be suitable value for the classification of pants or socks by pre-experiments. And the system was implemented using OpenNI, OpenCV, the VC + +.

Fig.4 shows a camera image used for background subtraction. Fig.5 shows a depth image that detected the floor area for background subtraction. After aplying initialize Kinect in 1-1 to 1-9, the system correctly cut out laundry pile field, recognized field, and small clothes field. Fig.6 shows a depth image that is divided into three floor area. Fig.7 shows binary image obtained from background subtraction. There is no object so fig.7 is just black. Fig.7 must compare with fig.10.

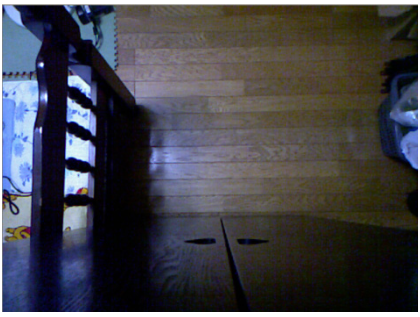


Fig. 4. Camera image

Fig. 5. Depth image

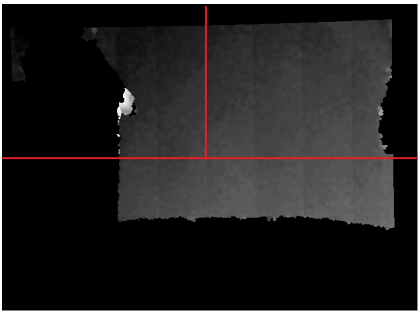
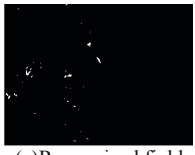
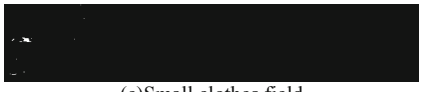


Fig. 6. Devided field image



(a)Recognized field

(b)Laundry pile field



(c)Small clothes field

Fig. 7. Binary image obtained from background subtraction

We showed the result of state determination and the transitions of state in Experiment in the following (Table2). Figures 8 to 10 are experimental example images of No.2 in Table2. Recognition of the IF part was able to meet the processing implemented from the experimental results.

Table 2. Results

No	Action	State recognition	Command
1	starting	4	(start)
2	place laundry pile	1	get out
3	socks→recognition field	2	socks pairing
4	socks→small clothes field	1	get out
5	pants→recognition field	2	pants tidying
6	pants→tidying	1	get out
7	socks→recognition field	2	socks pairing
8	socks→small clothes field	3	socks tidying
9	socks→tidying	3	socks tidying
10	socks→tidying	4	(finished)

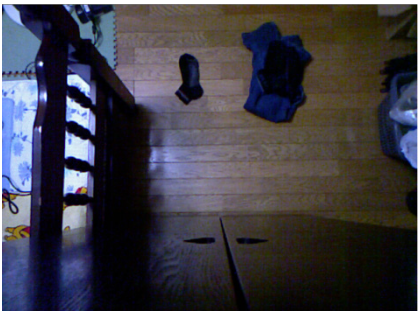


Fig. 8. Camera image of No.2 in Table2

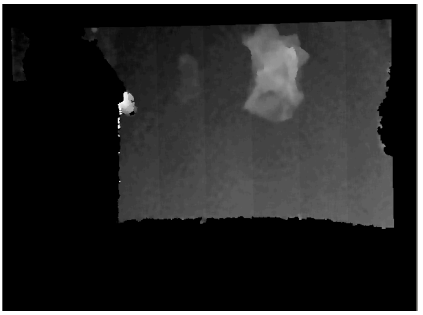


Fig. 9. Depth image of No.2 in Table2

(a)Recognized field

(b)Laundry pile field

(c)small clothes field

Fig.10. Binary image of No.2 in Table2 obtained from background subtraction

6 Conclusion

In this study, we proposed a laundry state determination system using depth information and camera image with Kinect. Results of the implementation experiment of the proposed method, it was possible to accurately determination the state of the laundry.

However, classification by only size must be refined, because type of clothes which need to discern are so many. Therefore, it is necessary to use characteristics such as shape to identify the type of more.

In addition, It is considered that the state determination would be able to do relatively easily because target laundry is limited to pants and socks. Case of assuming home use, it is not able to support at present. Therefore, we push forward with research so as to make accurate state transition by increasing the features to be used for recognition.

References

1. iRobot Corporation, http://www.irobot.com/us
2. Miyake, M., Manabe, Y., Uranishi, Y., Imura, M., Oshiro, O.: Voice Response System of Color and Pattern on Clothes for Visually Handicapped Person. In: 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4694–4697 (July 2013)
3. Zheng, X., Liu, N.: Color recognition of clothes based on k-means and mean shift. In: IEEE International Conference on Intelligent Control, Automatic Detection and High-End Equipment, pp. 49–53 (July 2012)
4. Wang, X., Li, K.: Pattern recognition based on fuzzy cluster for recognizing garment style in the photo. In: 9th International Conference on Computer-Aided Industrial Design and Conceptual Design, pp. 250–254 (November 2008)
5. Fabian, J., Young, T., Peyton Jones, J.C., Clayton, G.M.: Integrating the Microsoft Kinect With Simulink: Real-Time Object Tracking Example. IEEE/ASME Transactions on Mechatronics 19, 249–257 (2014)
6. Ramisa, A., Alenya, G., Moreno-Noguer, F., Torras, C.: Using depth and appearance features for informed robot grasping of highly wrinkled clothes. In: 2012 IEEE International Conference on Robotics and Automation (ICRA), pp. 1703–1708 (May 2012)
7. Willimon, B., Birchfield, S., Walker, I.: Classification of clothing using interactive perception. In: IEEE International Conference on Robotics and Automation, pp. 1862–1868 (2011)
8. Kragic, D., Björkman, M., Christensen, H.I., Eklundh, J.-O.: Vision for robotic object manipulation in domestic settings. Robotics and Autonomous Systems 52(1), 85–100 (2005)

9. Wang, P., Miller, S., Fritz, M., Darrell, T., Abbeel, P.: Perception for the Manipulation of Socks. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4877–4884 (2011)
10. Yamakazi, K., Inaba, M.: A cloth detection method based on image wrinkle feature for daily assistive robots. In: IAPR Conference on Machine Vision Applications, pp. 366–369 (2009)

Direction Control System on a Carrier Robot Using Fuzzy Logic Controller

Kevin Ananta Kurniawan, Darmawan Utomo, and Saptadi Nugroho

Satya Wacana Christian University
Jl.Diponegoro 52 - 60 Salatiga 50711, Jawa Tengah, Indonesia
kevin.ananta.kurniawan@gmail.com,
{du88,saptadi_nugroho}@yahoo.com

Abstract. In an autonomous mobile robot system the ability to control the robot manually is needed. For that reason a mechatronics system and control algorithm on carrier robot are designed and realized using Fuzzy Logic Controller. The carrier robot system which is designed consists of the robot mechatronics unit and the control center unit. The commands which are sent from the control center unit via local network received by the embedded system. These commands are forwarded to the microcontroller and translated into carrier robot's maneuver. An error correction algorithms using fuzzy logic controller is applied to regulate the actuator's speed. This fuzzy logic controller algorithm is implemented on embedded system which has a limitation on computational resources. The fuzzy controller gets its input from a rotary encoder and a magnetometer installed on the robot. The fuzzy logic controller algorithm using direction error protection has been able to detect and correct the direction error which the error value exceeds the predetermined threshold value (± 3 ° and ± 15 °). The carrier robot system has been able to run straight as far as 15 m with average deviation value of 22.2 cm. This fuzzy logic controller algorithm is able to give the response in the form of speed compensation value to maintain the direction of the carrier robot.

Keywords: fuzzy logic controller, fuzzy, embedded system, carrier robot, pulse width modulation.

1 Introduction

In this paper the design of the control system is realized on a two-wheeled carrier robot. The robot has two wheels which have the functions as actuators and a passive wheel on the back of the robot. The robot uses two wheels control system so that the space needed for the robot maneuver is smaller. The fuzzy logic algorithm implemented in the robot controller is used to correct the directional error of the robot when the robot moving in a straight line.

Several previous studies have discussed about the use of fuzzy algorithms for balance control on two wheels robot [1], to control two-wheel inverted pendulum [2],

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 27–36, 2015.
DOI: 10.1007/978-3-662-46742-8_3

cost control on a two-wheeled robot [3], and also the application of fuzzy algorithms for manipulator robot [4]. Based on these ideas, the authors are interested in researching and implementing fuzzy algorithm on embedded system for maintaining direction of a carrier robot.

The fuzzy controller algorithm designed in this paper is applied using Takagi-Sugeno-Kang (TSK) fuzzy inference models. Fuzzy system has two inputs such as an input of error angle and the value input of the actuator PWM. The output of the fuzzy controller is a crisp value representing the speed compensation value. The output will be added directly to the PWM of the motor which can be used to maintain the movement direction the robot.

The greater PWM will affect the higher output voltage and RPM. However the non linear systems (dynamic characteristics of the motor, the friction wheel, and the other) require complex models that must be known in advance if user wants to use PID technique. Fuzzy Logic Controller has several advantages such as having the ability to model non-linear functions and the ability to tolerate less accurate data.

2 Fuzzy Logic Controller

The fuzzy control system makes decisions using a language that can be understood by humans [5]. Knowledge possessed by humans can be easily applied in the design of a fuzzy control system. One of the most common ways of designing a fuzzy controller is to use a fuzzy rule-based system [6]. In Figure 1 the fuzzy controller architecture is shown in general. The fuzzification and defuzzification process are performed in the controller. The fuzzification process is a process to convert fuzzy variables to crisp variable, while the deffuzification process is a process to perform conversion of binary variables to fuzzy. The diagram block of a fuzzy logic controller is shown in figure 1.

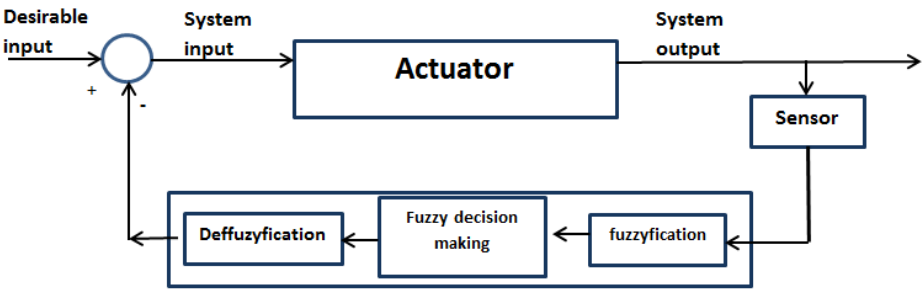


Fig. 1. Block diagram of the fuzzy logic controller

3 Takagi Sugeno Kang (TSK) Fuzzy Inference type [7]

In the TSK fuzzy inference model, the system outputs are linear constants and not fuzzy sets. The general rule of a TSK fuzzy inference model is:

$$\text{If input } \mathbf{1} = \boldsymbol{x_f} \text{ and input } \mathbf{2} = \boldsymbol{y_f} \text{ then output } \boldsymbol{z_f} = \boldsymbol{ax_f} + \boldsymbol{by_f} + \boldsymbol{c} \quad (1)$$

For the zero order TSK fuzzy model the value of $\boldsymbol{z_f}$ is a constant ($a = b = 0$). The value of z_i for each rule is weighted by the firing strength w_i of the rules. As example if the inputs were x and y then the firing strength is:

$$\boldsymbol{w_i} = \boldsymbol{And\ Method}\left(\boldsymbol{F_1}(\boldsymbol{x}), \boldsymbol{F_2}(\boldsymbol{y})\right) \quad (2)$$

Where $F_{1,2,\ldots,n}$ is the membership function of the input. The system final output is the weighted sum of each rules output.

$$\boldsymbol{Final\ Output} = \frac{\sum_{i=1}^{N} wi.zi}{\sum_{i=1}^{N} wi} \quad (3)$$

With N is the number of rules.

4 System Design

The carrier robot system consists of two main entities. They are main control and robot mechatronics. The Control Center has a graphical user interface (GUI) located on a Notebook PC. This GUI serves to command the robot and monitor the status of the carrier robot.

The robot mechatronics have two embedded modules. The modules are the Raspberry Pi and The Arduino microcontroller board. The Raspberry Pi serves as the medium level controller. Its function is to perform socket communication between the main control and the robot over a wireless local area network. The microcontroller serves as the low level controllers which read the sensor data and control the speed of the actuators. Figure 2 shows the designed system architecture of the carrier robot system.

5 Mechanical Design

The carrier robot mechanic has mechanical layers called physical layer which is a physical part of the robot. The mechanical layer consists of robot body, framework, motor, and wheels. The framework of the robot uses iron pipe material which has a thickness of 1.5 mm. Aluminum plate with a thickness of 1 mm is used to cover the body of the robot. The form of the mechanical robot design can be seen in Figure 3. The robot has two actuators that each actuator consists of a wheel which is connected to an electric DC motor. To maintain robot balance, there is one passive wheel on the back of the robot. The reason that the robot is designed with two wheels is to minimize the space needed for the robot's maneuver.

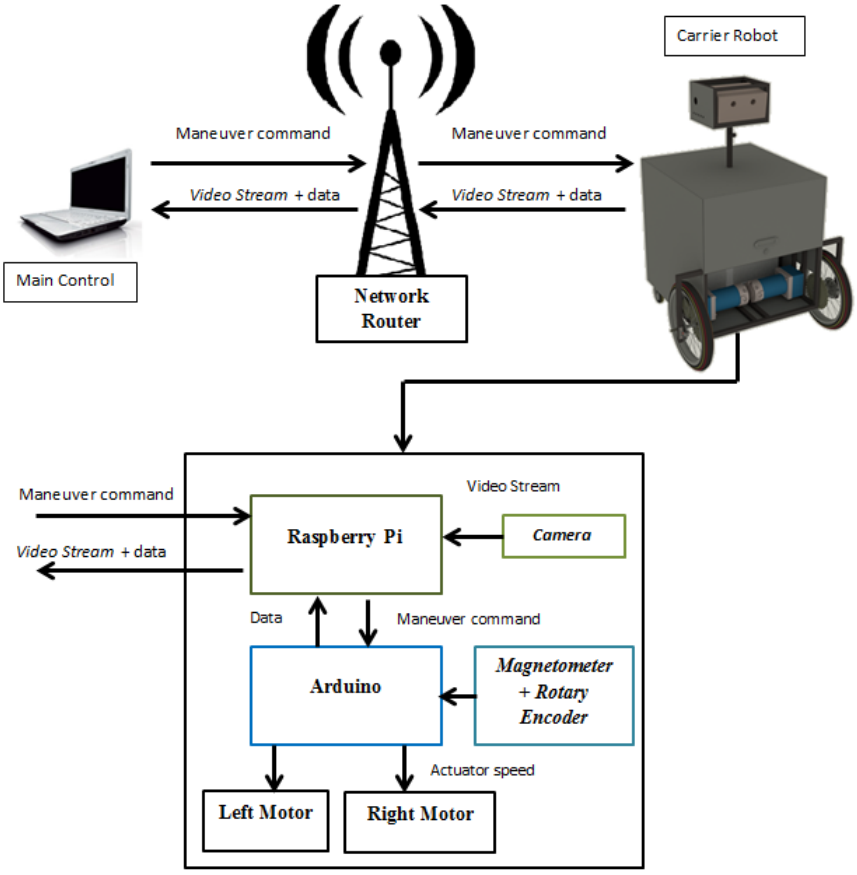


Fig. 2. Carrier Robot System Architecture

Fig. 3. Mechanical design of the carrier robot

6 Fuzzy Logic Controller Algorithm

Fuzzy Logic Controller Algorithm used in this system has two inputs and a single output. The two inputs are the error angle and PWM value that represents the speed of the motor. The output of the fuzzy algorithm is a speed compensation which its value will be added to the PWM value of each motor to regulate the speed of rotation of each wheel. The process of calculating the control algorithm has three main processes. The processes are the input fuzzyfication process, the fuzzy decision making process, and the defuzzyfication process. The membership degree of the input error angle and input PWM value is calculated in the input fuzzyfication process. An output of the fuzzy decision making process will be calculated from the membership value of the previous process. The five sets of fuzzy which are used to represent the error angle variable and PWM input can be seen in Figure 4.

Table 1. Fuzzy Rules of the left robot actuator

Error / PWM	Large Negative (LN)	Medium Negative (MN)	Normal (N)	Medium Positive (MP)	Large Positive (LP)
Very Slow (VS)	*LARGE POSITIVE*	*MEDIUM POSITIVE*	*SMALL*	*SMALL*	*SMALL*
Medium Slow (MS)	*LARGE POSITIVE*	*MEDIUM POSITIVE*	*SMALL*	*SMALL*	*MEDIUM NEGATIVE*
Normal (N)	*LARGE POSITIVE*	*MEDIUM POSITIVE*	*SMALL*	*MEDIUM NEGATIVE*	*MEDIUM NEGATIVE*
Medium Fast (MF)	*LARGE POSITIVE*	*MEDIUM POSITIVE*	*SMALL*	*MEDIUM NEGATIVE*	*LARGE NEGATIVE*
Very Fast (VF)	*LARGE POSITIVE*	*MEDIUM POSITIVE*	*SMALL*	*MEDIUM NEGATIVE*	*LARGE NEGATIVE*

Fuzzy decision-making process is a process of formulating the mapping from input to output using Fuzzy logic. The process of mapping from input to output is accomplished by determining a set of rules for every possible input combination. Table 1 shows the examples of rules that are applied to the left actuator.

The decision making is done by taking into account all the rules that its output is not zero. To obtain the membership degree of the output is calculated with the Fuzzy logic operator Min. In other words:

$$\mathbf{Uoutput = Min(uinput1, uinput2)} \quad (4)$$

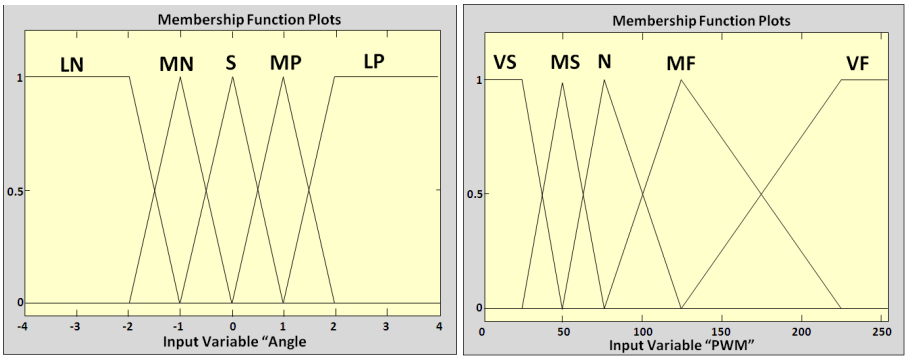
The output membership degree in Takagi-Sugeno-Kang (TSK) fuzzy method also known as the firing strength of a rule (wi). The firing strength of each output is calculated and then the deffuzyfication process would take place to agregate all the outputs to obtain a single crisp value.

The output speed offset constant is defined to map the fuzzy output to a crisp. The output equation of the TSK fuzzy method could be expressed as:

$$z_f = ax_f + by_f + c \tag{5}$$

Where z_f = output, x_f= input 1, input 2 and y_f = *a, b, c* constants. For TSK order 0, then the output is a constant C only. The output variable speed offset is divided into five fuzzy sets. The values of the constants for each output are: Large Negative = -15, Negative Medium = -7, Small = 0, Positive Medium = 10, Large Positive = 20.

As mentioned in equation (3), the final output of the system is the weighted sum of each rules output.



(a) (b)

Fig. 4. (a). Fuzzy sets for input variable angle error and (b) Fuzzy sets for input variable PWM

7 Experiments and Results

Several tests were performed for the fuzzy system implemented on the carrier robot. The first test is performed on the robot actuators to determine the value change of the revolution per minute (RPM) number for the given PWM values. The values obtained from the test are plotted on a graphic as shown in Figure 5. The values obtained from this test are used to determine fuzzy set boundaries for the PWM fuzzy sets.

The fuzzy logic algorithm implemented on the carrier robot is tested to know the characteristics of the error correction system. The test is performed by running the carrier robot while the microcontroller collects sensors data. This data is transmitted to the main control over the wireless network. The errors magnitude of the robot actuator is measured without any correction from the Fuzzy Logic Controller. The result of the error measurement is shown in Figure 6.

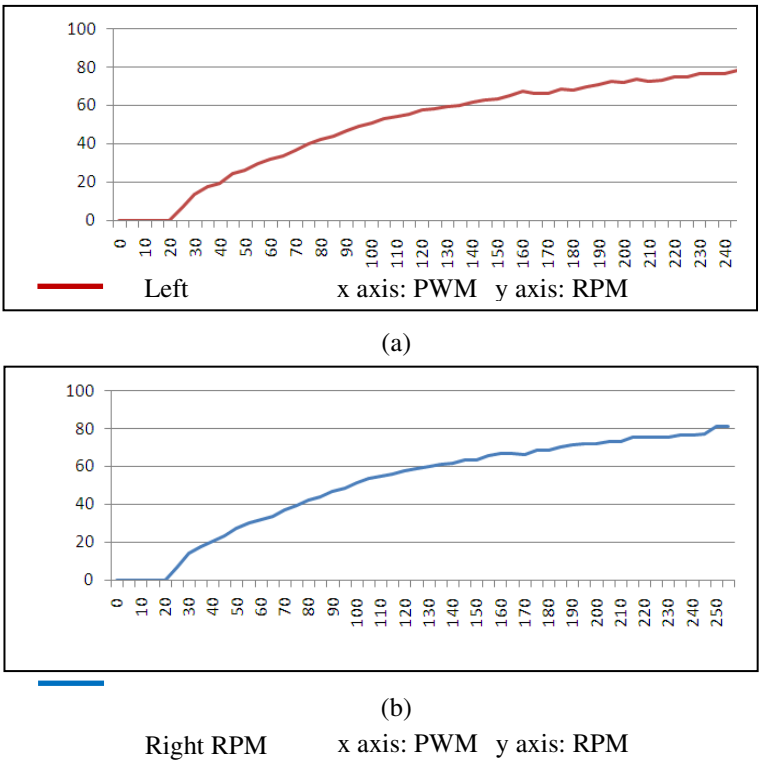


Fig. 5. The characteristics or the robot actuators (a) left; (b) right RPM

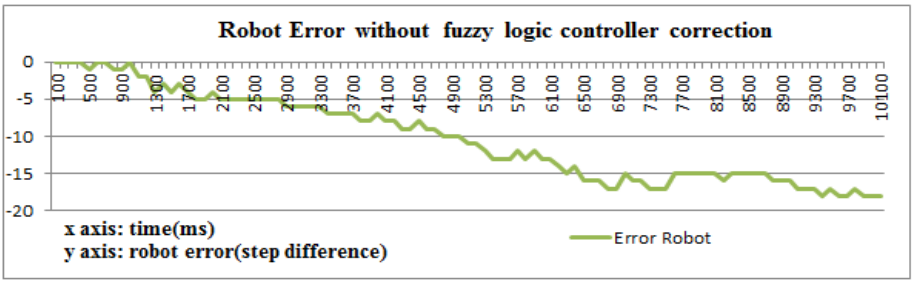


Fig. 6. The value of errors measurements without any correction from the Fuzzy Logic Controller

There are two fuzzy logic controller tests that are used to determine the error correction capability of each sensor used. The first test has an input of step difference value retrieved from rotary encoder only which the result of the test is shown in Figure 7. The second test has an input of direction angle difference value obtained from the magnetometer sensor only. The robot is run for 22 seconds while the data is collected by the microcontroller. The fuzzy logic controller algorithm that has input of step difference value retrieved from rotary encoder is able to correct the directional

error. The average of the error correction magnitude value obtained from this experiment is 0.54 steps in 22 seconds. The direction angle value can be changed due to the interference from the outside such as a skid which cannot be detected with the rotary encoder. To perform the experiment of fuzzy logic controller algorithm that has an input of angle difference value obtained from the magnetometer, the robot is run for 5 seconds and the input step difference value from the rotary encoder is not used. The fuzzy logic controller algorithm that has input of angle difference value is able to correct the directional error. The maximum deviation value of the robot direction in this experiment is 25°. This result can be seen in Figure 8. These maximum deviation values show that the robots will not run straight. The rotary encoder sensor and the magnetometer sensor can be used together to run the carrier robot straight and detect the direction angle of the carrier robot respectively.

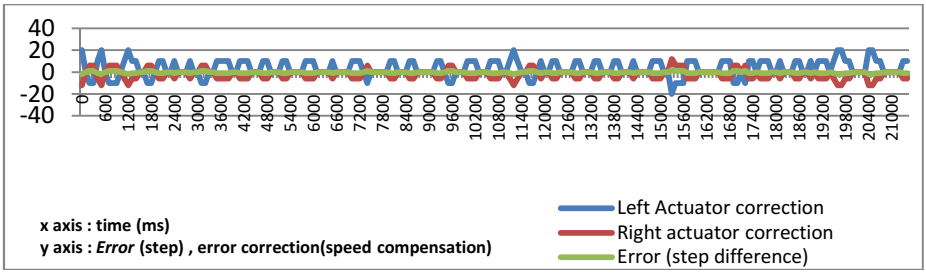


Fig. 7. The error correction tests with rotary encoder input only

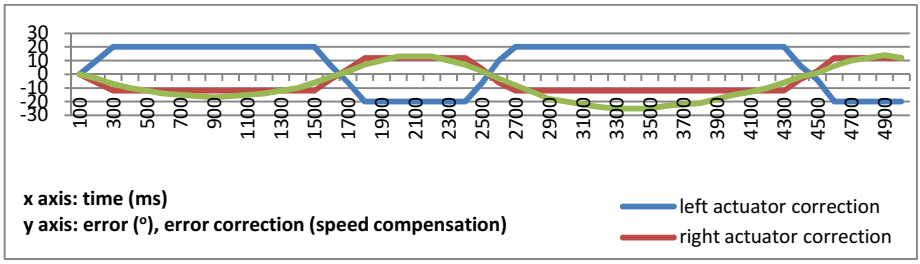


Fig. 8. The results of the error correction testing using a magnetometer input only

The error correction of the direction angle is performed using the fuzzy logic controller that has input from the magnetometer as seen in figure 9. The figure 9 shows the result of fuzzy logic controller error correction of the direction angle that has the upper threshold value of ± 15 °. If the error value of the direction angle exceeds the predetermined upper threshold value, then the magnetometer sensor is used for the input of fuzzy logic controller to perform the error correction. The Fuzzy Logic Controller will reuse sensor input rotary encoder to adjust the direction of the robot when the error value is lower than the predetermined lower threshold value ± 1.5°. The direction error protection implemented in the fuzzy logic controller algorithm is able to detect and correct large directional errors successfully.

The carrier robot is run in a straight line as far as 15 m for 15 times to determine the average deviation value. The test result of this experiment is shown in Table 2.

Table 2. Test result for the straight line forward command test

Test Results	Deviation (cm)
Average Deviation	**-9.2**
Magnitude of average deviation	**22.2**
Minimum	**0.0**
Maximum	**45.0**

In this experiment, the robot was able to run in a straight line as far as 15 m with an average deviation of 22 cm. It is necessary to perform a simulation which is used to know the characteristics of the control system for all possible input combination using MATLAB fuzzy toolbox.

The positive value of the error angle shows that the carrier robot will turn to the right, while the negative value of the error angle shows that the carrier robot will turn to the left. The positive error compensation is a positive value which will be added to the current motor PWM that is used to increase the speed of the actuator. The negative error compensation is a negative value which will be added to the current motor PWM that is used to reduce the speed of the actuator.

The output of the fuzzy logic controller algorithm is able to response the difference of speed compensation corresponding to the input of angle difference value successfully. The actual robot speed can influence the speed compensation value given by the fuzzy logic controller. If the value of the robot speed is increased then the value of the speed compensation will be increased.

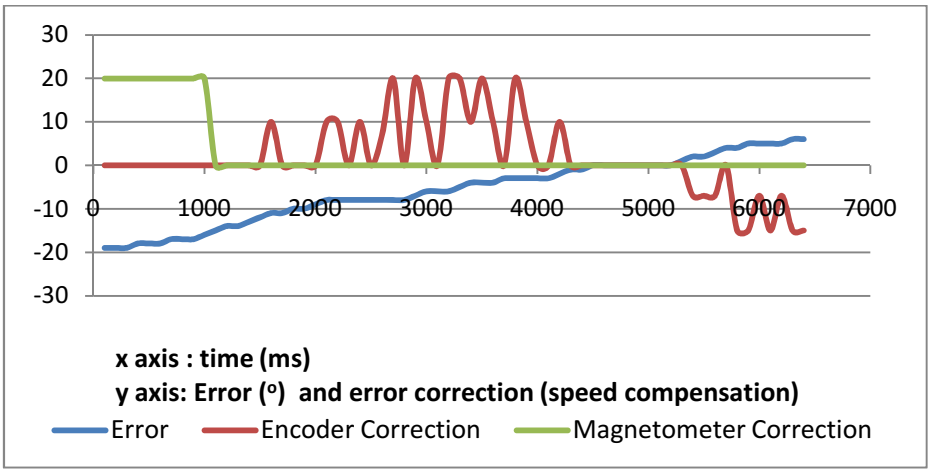


Fig. 9. Test result for error correction with direction error protection for error threshold value ± 15°

8 Conclusion

In this paper the design of the control system is realized on a two-wheeled carrier robot. Fuzzy logic algorithm implemented in the robot controller is used to correct the directional error of the robot while the robot runs in a straight line. The fuzzy logic controller algorithm used in this system has two inputs and a single output. The two inputs of the fuzzy logic controller algorithm are the direction angle error value and PWM value which represents the speed of the motor. The output of the fuzzy logic controller algorithm is a speed compensation which its value will be added to the PWM value of each motor. This speed compensation value is used to regulate the speed of each actuator. The fuzzy logic controller algorithm is able to detect and correct large directional errors. The robot was able to run in a straight line as far as 15 m with an average deviation of 22 cm. The output of the fuzzy logic controller algorithm is able to response the difference of speed compensation corresponding to the input of angle difference value.

References

1. Jian, X., et al.: Design and Implementation of a Takagi–Sugeno-Type Fuzzy Logic Controller on a Two-Wheeled Mobile Robot. IEEE Transactions on Industrial Electronics 60(12), 5717–5728
2. Huang, C.H., et al.: Design and Implementation of Fuzzy Control on a Two-Wheel Inverted Pendulum. IEEE Transactions on Industrial Electronics 58(7), 2988–3001
3. Sun, C.H., et al.: Switching T-S Fuzzy Model-Based Guaranteed Cost Control for Two-Wheeled Mobile Robots. International Journal of Innovative Computing, Information and Control ICIC International 8(5(A)) (May 2012)
4. Piltan, F., et al.: Design PID-Like Fuzzy Controller With Minimum Rule Base and Mathematical Proposed On-line Tunable Gain: Applied to Robot Manipulator. International Journal of Artificial Intelligence and Expert Systems (IJAE) 2(4) (2011)
5. Ross, T.J.: Fuzzy Logic with engineering application, 2nd edn. John WIlley& Sons Ltd., United Kingdom (2004)
6. Jamshidi, M., Vadiee, N., Ross, T.J.: Fuzzy Logic and control: Software and hardware Application, vol. 2. Prentice Hall International, New Jersey (1993)
7. Sugeno, M.: An Introductory Survey of Fuzzy Control. Inform. Sci. 36, 59–83 (1985)

Multidimensional Fuzzy Association Rules for Developing Decision Support System at Petra Christian University

Yulia, Siget Wibisono, and Rolly Intan

Petra Christian University, Surabaya, Indonesia
{yulia,rintan}@petra.ac.id, m26411074@john.petra.ac.id

Abstract. Academic records of student candidates and students of Petra Christian University (PCU) which have been stored so far have not been used to generate information. PCU's top-level management needs a way to generate information from the records. The generated information is expected to support the decision-making process of top-level management.

Before starting the application development, analysis and design of the student academic records and the needs of top-level management are done. The design stage produces a number of modeling that will be used to create the application.

The final result of the development is an application that can generate information using multidimensional fuzzy association rules.

Keywords: Application, Data Mining, Decision Support System, Multidimensional Fuzzy Association Rules.

1 Introduction

During this time, PCU has stored academic records of student candidates who enroll in PCU, such as math and english grades at their schools. In addition, after entering the university, PCU will save GPA of all students.

Academic records of student candidates and students that have been kept, have not been used to produce valuable information. PCU's top-level management needs a way to generate information from the records. The generated information is expected to support the decision-making process of top-level management.

With academic records of student candidates and students, information can be generated in the form of relationship between students' data using multidimensional fuzzy association rules. The students' data that can be used are schools, math, and english grade in their schools, specialization (science, social, literature, etc.), GPA, faculty, majors, gender, religion, and batch. Therefore, PCU need a software that can generate information needed by top-level management related to academic records of student candidates and students.

2 Data Mining

Data mining is one of the most important steps of the knowledge discovery in databases process. It is considered as significant subfield in knowledge management. Research

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 37–48, 2015.
DOI: 10.1007/978-3-662-46742-8_4

in data mining continues growing in business and in learning organization over coming decades[8]. Data mining is a process of extraction of useful information and patterns from huge data. It is also known as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis[9].

The development of Information Technology has generated great amount of databases and huge data in various areas. The research in databases and information technology has resulted in approach to store and manipulate this precious data for further decision making. The important reason that attracted many attentions in information technology and the discovery of meaningful information from large collections of data industry towards field of "Data mining" is due to the perception of "we are data rich but information poor". There is huge volume of data but we hardly able to generate them in to meaningful information and knowledge for decision making process in business[10].

Data mining derives its name from the similarities between finding valuable business information in a large database for example, finding linked products in gigabytes of store scanner data and mining a mountain for valuable ore. Both processes require either sifting through a great amount of material, and intelligently probing it to find exactly where the value resides. Given databases of sufficient size and quality, data mining technology can generate new business advantages and opportunities[10].

3 Multidimensional Association Rules

Association rule finds interesting association or correlation relationship among a large data set of items [1,2]. The discovery of interesting association rules can support decision making process.

Multidimensional association rules are association rules that involve two or more dimensions or predicates. Conceptually, a multidimensional association rule, $A \Rightarrow B$ consists of A and B as two datasets, called premise and conclusion, respectively.

Formally, A is a dataset consisting of several distinct data, where each data value in A is taken from a distinct domain attribute in D as given by

$$A = \{a_j \mid a_j \in D_j, \text{ for some } j \in \mathrm{N}_n\},$$

where, $D_A \subseteq D$ is a set of domain attributes in which all data values of *A* come from.
Similarly,

$$B = \{b_j \mid b_j \in D_j, \text{ for some } j \in \mathrm{N}_n\},$$

where, $D_B \subseteq D$ is a set of domain attributes in which all data values of *B* come from.

For example, database of medical track record patients is analyzed for finding association (correlation) among diseases taken from the data of complicated several diseases suffered by patients in a certain time. Additional related information regarding the identity of patients, such as *age*, *occupation*, *sex*, *address*, *blood type*, etc., may have a correlation to the illness of patients. Considering each data attribute as a predicate, it can therefore be interesting to mine association rules containing *multiple* predicates, such as:

Rule-1:

$Age(X,"60") \wedge Smk(X,"yes") \Rightarrow Dis(X,"LungCancer")$,

where there are three predicates, namely *Age*, *Smk* (*smoking)* and *Dis (disease)*. Association rules that involve two or more dimensions or predicates can be referred to as multidimensional association rules.

From Rule-1, it can be found that A={60, *yes*}, B={Lung Cancer}, D_A={*age*, *smoking*} and D_B={*disease*}.

Considering $A \Rightarrow B$ is an interdimension association rule, it can be proved that $|D_A| = |A|$, $|D_B| = |B|$ and $D_A \cap D_B = \varnothing$.

Support of A is then defined by:

$$\text{supp}(A) = \frac{|\{t_i \mid d_{ij} = a_j, \forall a_j \in A\}|}{r} \tag{1}$$

where r is the number of records or tuples (see Table 1, r=12).

Alternatively, r in (1) may be changed to $|Q(D_A)|$ by assuming that records or tuples, involved in the process of mining association rules are records in which data values of a certain set of domain attributes, D_A, are not null data. Hence, (1) can be also defined by:

$$\text{supp}(A) = \frac{|\{t_i \mid d_{ij} = a_j, \forall a_j \in A\}|}{|Q(D_A)|} \tag{2}$$

where $Q(D_A)$, simply called *qualified data* of D_A, is defined as a set of record numbers (t_i) in which all data values of domain attributes in D_A are not null data. Formally, $Q(D_A)$ is defined as follows.

$$Q(D_A) = \{t_i \mid d_{ij} \neq null, \forall D_j \in D_A\} \tag{3}$$

Similarly,

$$\text{supp}(B) = \frac{|\{t_i \mid d_{ij} = b_j, \forall b_j \in B\}|}{|Q(D_B)|} \tag{4}$$

Similarly, $\text{support}(A \Rightarrow B)$ is given by

$$\begin{aligned}\text{supp}(A \Rightarrow B) &= \text{supp}(A \cup B) \\ &= \frac{|\{t_i \mid d_{ij} = c_j, \forall c_j \in A \cup B\}|}{|Q(D_A \cup D_B)|}\end{aligned} \tag{5}$$

where $Q(D_A \cup D_B) = \{t_i \mid d_{ij} \neq null, \forall D_j \in D_A \cup D_B\}$ $\text{conf}(A \Rightarrow B)$ as a measure of certainty to assess the validity of $A \Rightarrow B$ is calculated by

$$\text{conf}(A \Rightarrow B) = \frac{|\{t_i \mid d_{ij} = c_j, \forall c_j \in A \cup B\}|}{|\{t_i \mid d_{ij} = a_j, \forall a_j \in A\}|} \tag{6}$$

A and *B* in the previous discussion are datasets in which each element of *A* and *B* is an atomic crisp value. To provide a generalized multidimensional association rules, instead of an atomic crisp value, we may consider each element of the datasets to be a dataset of a certain domain attribute. Hence, *A* and *B* are sets of set of data values or sets of datasets. For example, the rule may be represented by

Rule-2:

$age(X, \text{"20...60"}) \wedge smoking(X, \text{"yes"}) \Rightarrow$

$disease(X, \text{"bronchitis, lung cancer"})$,

where A={{20…60}, {yes}} and B={{bronchitis, lung cancer}}.

Simply, let *A* be a generalized dataset. Formally, *A* is given by

$$A = \{A_j \mid A_j \subseteq D_j, \text{ for some } j \in \mathrm{N}_n\}.$$

Corresponding to (2), support of *A* is then defined by:

$$\text{supp}(A) = \frac{|\{t_i \mid d_{ij} \subseteq A_j, \forall A_j \in A\}|}{|Q(D_A)|} \tag{7}$$

Similar to (5),

$$\begin{aligned}\text{supp}(A \Rightarrow B) &= \text{supp}(A \cup B) \\ &= \frac{|\{t_i \mid d_{ij} \subseteq C_j, \forall C_j \in A \cup B\}|}{|Q(D_A \cup D_B)|}\end{aligned} \tag{8}$$

Finally, $\text{conf}(A \Rightarrow B)$ is defined by

$$\text{conf}(A \Rightarrow B) = \frac{|\{t_i \mid d_{ij} \subseteq C_j, \forall C_j \in A \cup B\}|}{|\{t_i \mid d_{ij} \subseteq A_j, \forall A_j \in A\}|} \tag{9}$$

To provide a more meaningful association rule, it is necessary to utilize *fuzzy sets* over a given database attribute called *fuzzy association rule* as discussed in [4,5].

Formally, given a crisp domain D, any arbitrary fuzzy set (say, fuzzy set A) is defined by a membership function of the form [2,3]:

$$A : D \longrightarrow [0,1]. \tag{10}$$

To provide a more generalized multidimensional association rules, we may consider A and B as sets of fuzzy labels[6]. Simply, A and B are called fuzzy datasets. Rule-2 is an example of such rules, where A={*young*, *yes*} and B={*bronchitis*}. Here *young*, *yes* and *bronchitis* are considered as fuzzy lables. A fuzzy dataset is a set of fuzzy lables/ data consisting of several distinct fuzzy labels, where each fuzzy label is represented by a fuzzy set on a certain domain attribute. Let A be a fuzzy dataset. Formally, A is given by

$$A = \{A_j \mid A_j \in \mathrm{F}(D_j), \text{ for some } j \in \mathrm{N}_n\},$$

where $\mathrm{F}(D_j)$ is a fuzzy power set of D_j, or in other words, A_j is a fuzzy set on D_j.

Corresponding to (7), support of A is then defined by:

$$\mathrm{supp}(A) = \frac{\sum_{i=1}^{r} \inf_{A_j \in A} \{\mu_{A_j}(d_{ij})\}}{|Q(D_A)|} \tag{11}$$

Similar to (5),

$$\begin{aligned} \mathrm{supp}(A \Rightarrow B) &= \mathrm{supp}(A \cup B) \\ &= \frac{\sum_{i=1}^{r} \inf_{C_j \in A \cup B} \{\mu_{C_j}(d_{ij})\}}{|Q(D_A \cup D_B)|} \end{aligned} \tag{12}$$

$\mathrm{conf}(A \Rightarrow B)$ is defined by

$$\mathrm{conf}(A \Rightarrow B) = \frac{\sum_{i=1}^{r} \inf_{C_j \in A \cup B} \{\mu_{C_j}(d_{ij})\}}{\sum_{i=1}^{r} \inf_{A_j \in A} \{\mu_{A_j}(d_{ij})\}} \tag{13}$$

The correlation betwen two fuzzy datasets can be defined by the following definition.

$$\mathrm{corr}(A \Rightarrow B) = \frac{\sum_{i=1}^{r} \inf_{C_j \in A \cup B} \{\mu_{C_j}(d_{ij})\}}{\sum_{i=1}^{r} \inf_{A_j \in A} \{\mu_{A_j}(d_{ij})\} \times \inf_{B_k \in B} \{\mu_{B_k}(d_{ik})\}} \tag{14}$$

4 Research Methodology

4.1 Problems Analysis

There are several problems faced by PCU, such as:

1. PCU's top-level management takes decisions for the promotion or cooperation purpose based solely on estimates and habits, has not taken advantage of the existing academic records.
2. PCU's Faculties/Majors Promotion Team has not equipped with information or facts about the academic condition of PCU's students while promoting faculties/majors to high schools.
3. There is no feature in the current academic information system that can show the relationship between students' data.

4.2 Requirements Analysis

From the problems listed above, it can be concluded that the PCU's top-level management requires a computer-based system to assist in generating PCU's students academic records, that is a data mining-based information systems that can produce association rules of students' attributes. This system obtains data from the ETL process and has a multidimensional concept that shows the relationships between students' attributes. The dimensions used are schools, math and english grade in their schools, specialization (science, social, literature, etc.), GPA, faculty, majors, gender, religion, and batch.

4.3 Extract, Transform, and Load

Extract, Transform, and Load (ETL) is a function that integrates data and involves extracting data from sources, transforming it to be more valid, and loading it into a data warehouse[7]. This process begins by importing the data from the database. The imported data is religions, majors, schools, specializations, student candidates, students, and student admissions. Next, the imported data is transformed into more valid data and loaded into data warehouse.

4.4 Determination of Fuzzy Values

Determination of fuzzy values is done by establishing a group fuzzy set. First, user must input the name and choose the attribute, such as religions, majors, schools, GPA, math grade, etc. Next, user can make as many fuzzy sets as he/she wants inside the group fuzzy set made. User need to fill the name and the description of the fuzzy set. There are two types of fuzzy set based on the attribute of the group fuzzy set, numerical and non-numerical. For numerical, user can input as many points as he/she wants to form fuzzy membership function. A point includes crisp value and the membership degree of the crisp value to the fuzzy set. For non-numerical, user must input membership degree for every members of the attribute. Flowchart for determination of fuzzy values can be seen on Figure 1.

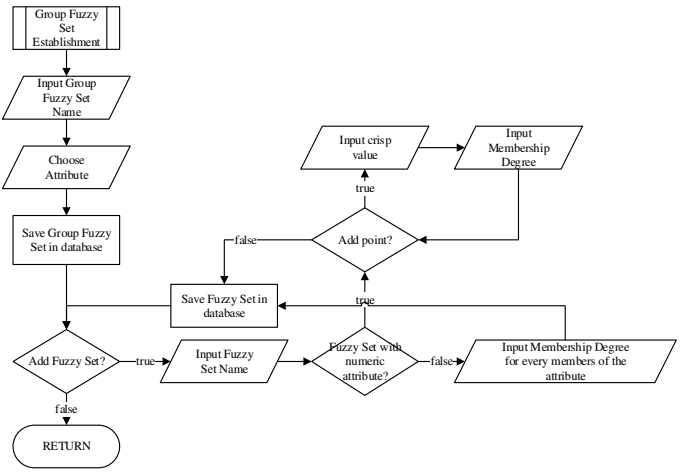


Fig. 1. Flowchart for Determination of Fuzzy Values

4.5 Customization of Fuzzy Association Rules

Customization of fuzzy association rules is done to generate fuzzy association rules report to support the decision-making process of top-level management. First, user must input the name and choose the attributes that will be used to generate the rules. After choosing the attributes, user must choose the group fuzzy set(s) of the attributes. Next, the application will generate the rules and save them in database. The user can see the whole report and filter the rules based on the support, confidence, and correlation value of the rules. Flowchart for customization of fuzzy association rules can be seen on Figure 2.

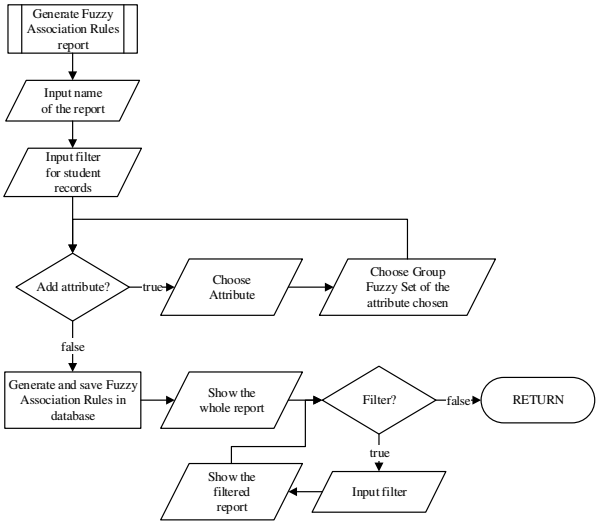


Fig. 2. Flowchart for Fuzzy Association Rules

5 Results

A test is conducted to prove the accuracy of the developed application to calculate support, confidence, and correlation of the multidimensional fuzzy association rules generated. The test is started from a given simple academic records of students with three attributes, such as major, math grade, Grade Point Average (GPA) as shown in Table 1.

Table 1. Academic Records of Students

Student	Major	Math grade	GPA
1	English Literature	74	3.34
2	Civil Engineering	75	3.41
3	Civil Engineering	90	3.9
4	Interior Design	86	3.75
5	Interior Design	78	3.45
6	Business Management	76	3.23
7	Business Management	68	3.35
8	Business Management	89	3.56
9	Informatics Engineering	91	3.84
10	Informatics Engineering	71	3.01
11	Science Communication	79	2.71
12	Science Communication	76	3.03

The test is conducted using three attributes, such as major, math grade, and GPA. First, we must determine how to convert each crisp value into fuzzy value for every attributes. Major is a non-numerical attribute, so we must determine the fuzzy value for each major. For example, we make a group fuzzy set for major named 2014 which has a fuzzy set named Engineering. Inside this fuzzy set, we determine business management has a membership degree of 0.2, civil engineering has a membership degree of 1, and so on as shown on Figure 3.

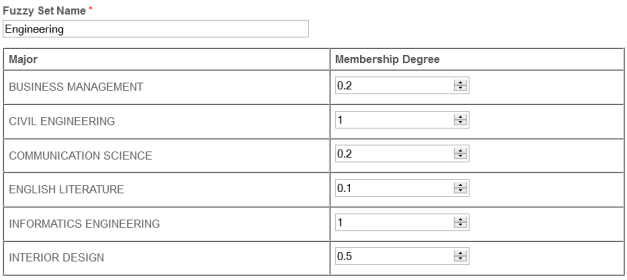


Fig. 3. Input for Major Fuzzy Values

Math grade is a numerical attribute, so that the fuzzy value of math grade will be calculated through a fuzzy membership function which is formed from the points stored in the fuzzy set. For example, we make a group fuzzy set for math grade named 2014 which has a fuzzy set named High. Inside this fuzzy set, we determine math grade of 0 has a membership degree of 0, math grade of 75 has a membership degree of 0, math grade of 95 has a membership degree of 1, and math grade of 100 has a membership degree of 1 as shown on Figure 4.

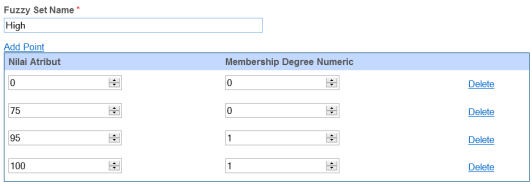


Fig. 4. Input for Math Grade Fuzzy Membership Function

These four points will form fuzzy membership function as shown on Figure 5.

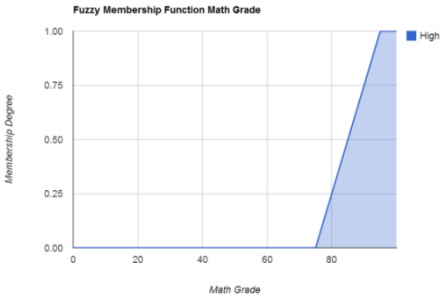


Fig. 5. Visualization of Math Grade Fuzzy Membership Function

For example, if a student has math grade of 90, then the application will look for its membership degree through the equation of the line that formed the point (75, 0) and (95, 1), as given by $y = \{ x \mid 0.05x - 3.75, \text{ for } 75 \leq x \leq 95\}$. Thus, membership degree of 90 is 0.05 * 90 - 3.75 = 0.75.

GPA is a numerical attribute like the math grade, so that the fuzzy value of GPA will also be calculated through a fuzzy membership function. For example, we make a group fuzzy set for GPA named 2014 which has a fuzzy set named High. Inside this fuzzy set, we determine points, such as (0, 0), (3.2, 0), (3.4, 0.6), (3.7, 1), and (4, 1) as shown on Figure 6.

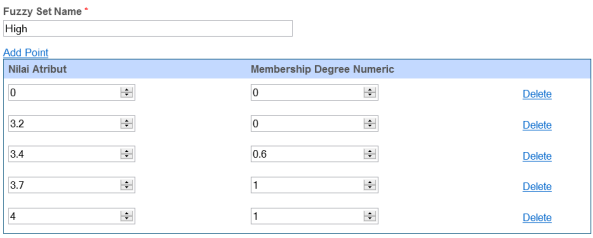


Fig. 6. Input for GPA Fuzzy Membership Function

This example of engineering fuzzy set for major attribute, high math grade fuzzy membership function, and high GPA fuzzy membership function are determined by interviewing one of PCU's structural officers. Next, we choose the attributes that are used during this test and each attribute's group fuzzy set that we just made before as shown on Figure 7.

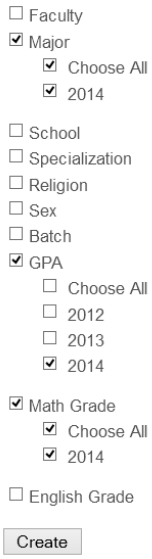


Fig. 7. Input Attributes for Fuzzy Association Rules

This test will generate all combinations of fuzzy association rules using every fuzzy sets of the attributes chosen. For example, one of the rules may be represented by:

Rule-3:

$$Major(X, \text{"Engineering"}) \wedge Math(X, \text{"high"}) \Rightarrow GPA(X, \text{"high"})$$

Rule-3 is a fuzzy rule, where A={Engineering, high} and B={high}. Next, each academic records of students shown in Table 1 will be converted using the fuzzy sets to fuzzy values as shown in Table 2.

Table 2. Calculation of Fuzzy Values

	α	β	ϒ	X	Y	X*Y	Z
1	0.1	0	0.42	0	0.42	0	0
2	1	0	0.613	0	0.613	0	0
3	1	0.75	1	0.75	1	0.75	0.75
4	0.5	0.55	1	0.5	1	0.5	0.5
5	0.5	0.15	0.667	0.15	0.667	0.10005	0.15
6	0.2	0.05	0.09	0.05	0.09	0.0045	0.05
7	0.2	0	0.45	0	0.45	0	0
8	0.2	0.7	0.813	0.2	0.813	0.1626	0.2
9	1	0.8	1	0.8	1	0.8	0.8
10	1	0	0	0	0	0	0
11	0.2	0.2	0	0.2	0	0	0
12	0.2	0.05	0	0.05	0	0	0
Σ	6.1	3.25	6.053	2.7	6.053	2.31715	2.45

Note:
$\alpha = \mu_{engineering}$ (major)
$\beta = \mu_{high}$ (math)
$\gamma = \mu_{high}$ (GPA)
$X = \min(\alpha,\beta)$
$Y = \min(\gamma)$
$Z = \min(\alpha,\beta,\gamma)$

Therefore, support of Rule-3 can be calculated by (12),
supp(Rule-3) = 2.45 / 12 = 0.20417

On the other hand, confidence of Rule-3 can be calculated by (13),
conf(Rule-3) = 2.45 / 2.7 = 0.90741

On the other hand, correlation of Rule-3 can be calculated by (14),
corr(Rule-3) = 2.45 / 2.31715 = 1.05733

The manually calculated support, confidence, and correlation of Rule-3 are match with the output of the fuzzy association rules generated by this test as shown on Figure 8.

Fuzzy Association Rules with 3 Attributes				Support	Confidence	Correlation
Major = Engineering (2014)	Math Grade = High (2014)	==>	GPA = High (2014)	0.20417	0.90741	1.05733

Fig. 8. Example Output of Fuzzy Association Rules

To evaluate this application, research on the use of this application is conducted. Samples of this research is five structural officers of PCU. To collect the data, distributed a questionnaire containing indicators to evaluate the use of the application. The indicators include display of application, determination of fuzzy values, customization of fuzzy association rules, ease of use, the ability to address the needs of users, and overall. From the data collected, the calculation of the percentage of user satisfaction in using this application is done.

Assessment of the feasibility of the application:
1. Display of application is 100% good
2. Determination of fuzzy values is 80% good
3. Customization of fuzzy association rules is 80% good
4. Ease of use is 100% good
5. The ability to address the needs of users is 60% good
6. Overall is 100% good

6 Conclusion

The generated fuzzy association rules have been tested and matched with the Multidimensional Fuzzy Association Rules algorithm and the reality of academic situation of PCU's students. From the assessment, obtained that overall application is 100% good. This suggests that the application developed has benefits for PCU and can be continued for the purpose of decision-making process by top-level management.

References

1. Han, J., Kamber, M., Pei, J.: Data mining: Concepts and Techniques, 3rd edn. Morgan Kaufmann, San Fransisco (2012)
2. Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, New Jersey (1995)
3. Zadeh, L.A.: Fuzzy Sets and Systems. International Journal of General Systems 17, 129–138 (1990)
4. Intan, R.: A Proposal of Fuzzy Multidimensional Association Rules. Jurnal Informatika 7(2) (2006)
5. Intan, R.: A Proposal of an Algorithm for Generating Fuzzy Association Rule Mining in Market Basket Analysis. In: Proceeding of CIRAS (IEEE), Singapore (2005)
6. Intan, R.: Hybrid-Multidimensional Fuzzy Association Rules from a Normalized Database. In: Proceedings of International Conference on Convergence and Hybrid Information Technology. IEEE Computer Society, Daejeon (2008)
7. Gour, V., Sarangdevot, S.S., Tanwar, G.S., Sharma, A.: Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse. International Journal on Computer Science and Engineering 2(3), 786–789 (2010)
8. Silwattananusarn, T., Tuamsuk, K.: Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012. International Journal of Data Mining & Knowledge Management Process (IJDKP) 2(5), 13–24 (2012)
9. Ramageri, B.M.: Data Mining Techniques and Applications. Indian Journal of Computer Science and Engineering 1(4), 301–305 (2010)
10. Padhy, N., Mishra, P., Panigrahi, R.: The Survey of Data Mining Applications and Feature Scope. International Journal of Computer Science, Engineering and Information Technology (IJCSEIT) 2(3), 43–58 (2012)

Genetic Algorithm and Heuristic Approaches

Genetic Algorithm for Scheduling Courses

Gregorius Satia Budhi, Kartika Gunadi, and Denny Alexander Wibowo

Petra Christian University, Informatics Department
Siwalankerto 121-131, Surabaya, East Java, Indonesia
{greg,kgunadi}@petra.ac.id

Abstract. In the university, college students must be register for their classes. But there still many college student that was confused on how to make a good classes schedule for themselves. Mainly because of many variables and considerations to be made, for examples, they have to consider how hard the classes they are going to take, and also, they still have to consider their exam schedules and also their availability time as well. Genetic Algorithm is one of many methods that can be used to create a schedule. This method determines the best schedule using fitness cost calculation which can compare the quality of one schedule against the other. Then, using crossover, mutation, and elitism selections, we can determine better schedules. Based on the result of the survey held before, 70% of the respondents gave point 4 and 30% of the respondents gave point 5 out of 5 for the quality of the schedule made using this applications.

Keywords: Artificial Intelligence, Genetic Algorithm, Courses Schedule.

1 Introduction

Schedule is very important to manage an activity, especially when the activity was carried out in a large organization and held for long-term time or routine.Within the scope of the university, the course schedule also is critical There are still students who are confused when creating their own class schedules, because a lot of consideration, for example: what subjects can be taken without clashing with another schedule, a student who wants to the day - certain the day No lectures, due to personal reasons or other student affairs, student who wants to take courses that would like to correspond with their fields of specialization, and so forth. These things make them have trouble making a personal course schedule. Through these problems, we conduct research on a system that can help perform automatic scheduling lectures for students by the day and the course desired by the student. The method used for the manufacture of automatic schedule is Genetic Algorithm (GA). The reason is because GA is a method that can find a solution that is both complex problems.

2 Genetic Algorithm

Genetic Algorithm is a method to search an optimal solution for a problem. The method will find a good solution by crossovers a possible solution with another solution

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 51–63, 2015.
DOI: 10.1007/978-3-662-46742-8_5

to create new solutions. After that method will mutate the new solutions so that they are have parts of solution from the parents but not really same with the parent. The process begin with the creation of random population of valid solutions / chromosomes, then GA will count the fitness costs of each chromosomes in the population. After that two chromosomes will be chose to crossover and produce offspring. Then the offspring will be mutated. This process will be repeated till the stop condition was reached [1, 2].

2.1 Selection Criteria

The commonly used chromosome selection technique is the roulette wheel [3, 4]. It is like spinning a roulette wheel where each chromosome has a segment on the wheel proportional to its fitness. The roulette wheel is spun, and when the arrow comes to rest on one of the segments, the corresponding chromosome is selected. Each chromosome is given a slice of a circular roulette wheel. The most fits chromosomes occupy the largest areas, whereas the least fits have much smaller segments in the roulette wheel. The area of the slice within the wheel is equal to the chromosome fitness ratio. To select a chromosome for mating, a random number is generated in the interval (0, 100), and the chromosome whose segment spans the random number is selected [2].

2.2 Reproduction

There are several methods used in the GA Reproduction one of them is crossover. Crossover is the crosses of at least 2 solutions, where through parts crossed, will generate new solutions. Through this crosses GA will add solutions and variations that are useful later in the scoring stage when searching for a better solution [5].

Besides the crossover, there is another way to complete reproduction, namely mutation. Mutation is a method that converts a portion of the solution. Although altered a little bit, it can create different variations resulting in different solutions and even new solution. One method used in this mutation is shuffle. [5]. Mutation role is to provide a guarantee that the search algorithm is not trapped on a local optimum [6].

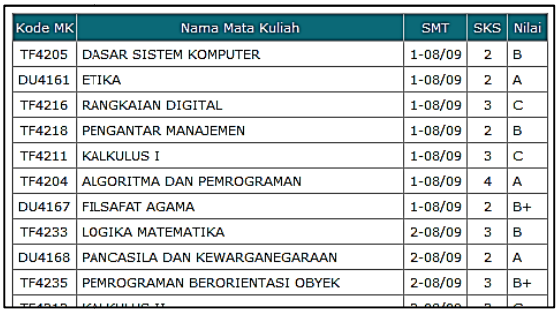
3 Courses Enrolling on Study Plan Registration Time

Students will enroll subjects which wants to be taken at the beginning of the semester, during the study plan registration (SPR/PRS). PRS periods are divided into three times, namely the PRS-1, PRS-2 and PRS-3. During the PRS-1, after receiving Study Results Card (SRC/KHS), and get permission to register from faculty trustee, students can enroll subjects that have been opened by the department. For example, the list of courses that are opened by PCU Informatics department could be seen in Figure 1.

HARI	JAM	SAMPAI	NOVAK	NAMA	KLS
SENIN	07.30	10.30	TF4303	APLIKASI SISTEM PAKAR	
	08.30	11.30	TF4343	TEKNOLOGI WEB	B
	09.30	11.30	TF4205	DASAR SISTEM KOMPUTER	A
	09.30	11.30	TF4259	METODOLOGI PENELITIAN	B
	09.30	11.30	TF4273	BASIS DATA LANJUTAN	C
	13.30	16.30	TF4220	PENG. MANAJEMEN & BISNIS	
	13.30	15.30	TF4211	KALKULUS I	B
	13.30	16.30	TF4225	PAKET PROGRAM APLIKASI	C
	14.30	17.30	TF4241	SISTEM INFORMASI	A
	14.30	17.30	TF4413	E - BUSINESS	B
	15.30	18.30	TF4361	PEMROGRAMAN APLIKASI JAVA	
	15.30	18.30	TF4363	PEMROGRAMAN OPEN SOURCE	
	15.30	18.30	TF4337	SISTEM LOGIKA FUZZY	

Fig. 1. The list of courses that are opened for one semester

Then in Figure 2 you could see the example of courses that have been taken by students that called the transcript.



Kode MK	Nama Mata Kuliah	SMT	SKS	Nilai
TF4205	DASAR SISTEM KOMPUTER	1-08/09	2	B
DU4161	ETIKA	1-08/09	2	A
TF4216	RANGKAIAN DIGITAL	1-08/09	3	C
TF4218	PENGANTAR MANAJEMEN	1-08/09	2	B
TF4211	KALKULUS I	1-08/09	3	C
TF4204	ALGORITMA DAN PEMROGRAMAN	1-08/09	4	A
DU4167	FILSAFAT AGAMA	1-08/09	2	B+
TF4233	LOGIKA MATEMATIKA	2-08/09	3	B
DU4168	PANCASILA DAN KEWARGANEGARAAN	2-08/09	2	A
TF4235	PEMROGRAMAN BERORIENTASI OBYEK	2-08/09	3	B+
TF4212	KALKULUS II	2-08/09	3	C

Fig. 2. Student transcript example

Based on the example of courses offered (Figure 1) and student transcript (Figure 2), the list of courses that can be taken by the student in the next semester could be seen in Table 1.

Table 1. The example of courses which could be taken by student base on Figure 1 and 2

Course Code and Name	Class	Type	Prerequisite
TF4251 – Komunikasi Data	A,B,C	Mandatory	K. RD
TF4219 – Struktur Data	A	Mandatory	K. PBO
TF4229 – Basis Data	A	Mandatory	-

For courses with prerequisites "Had attended", if the student has taken the prerequisite courses but did not passed, he could take courses with these prerequisites. But for courses with prerequisites "Should be passed", the courses prerequisites should be passed before the desired courses taken. Courses who had been taken by a student, still could be taken again even after passed, with the aim to improve the credit point.

For courses with 'Elective' type, it can be taken if it meets prerequisites and the minimum number of credits that have been completed by the student is 60 credits. For courses with 'Mandatory' type, it could be taken when the student have met the prerequisites of the courses.

When registering on the PRS, students are given choice of priority 1 and 2. For example, if a student has a value of 20 for priority 1, then the student should only enroll 20 Semester Credit Units (SCU/SKS) with priority 1, and the others are priority 2. The priority 1 is very important because the chance it is likely to be accepted in the class larger than the one with priority 2.

During the PRS-2 students could revise or add more courses to be taken up to a maximum limit of credits. In addition, students could cancel courses which has received in PRS-1. While the PRS-3 is the time to cancel courses that have been received in the PRS-1 and PRS-2.

4 Genetic Algorithm Design

4.1 Chromosome Design

Chromosome is an important factor in the genetic algorithm, in which the design of chromosomes will affect how the crossover implemented. Form of chromosome could be seen in Figure 3.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Day 0	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Day 1	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Day 2	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Day 3	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Day 4	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Day 5	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

Fig. 3. Chromosome Design

The chromosome has day index starting from zero to the fifth indexes. Where in each day they have hour index starting from zero to the fourteen indexes.

Day index only have six days, because on the day of Sunday is certainly no lectures. As for the hour index in each day, containing fourteen ranging from 07.30 to 22.30 with a time range of 1 hour. For example of chromosomes that have been filled could be seen in Figure 4.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Day 0	[2]	[2]	[2]	[0]	[0]	[6]	[6]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Day 1	[0]	[0]	[49]	[49]	[0]	[0]	[0]	[0]	[10]	[10]	[0]	[0]	[0]	[0]	[0]
Day 2	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Day 3	[0]	[0]	[0]	[0]	[0]	[0]	[19]	[19]	[19]	[0]	[0]	[0]	[0]	[0]	[0]
Day 4	[0]	[33]	[33]	[33]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]
Day 5	[0]	[0]	[0]	[0]	[44]	[44]	[44]	[0]	[0]	[0]	[0]	[0]	[0]	[0]	[0]

Fig. 4. Example of chromosomes that have been filled

The values present in a chromosome are the id of courses that opened in a semester. Complete data from the ids are stored in a database table with the design as shown in Table 2.

Table 2. Id Class information

Description	Sample Values
ID class	49
Course name	Struktur Data
Class	A
SKSk	2
SKSp	1
SKSs	0
SKSr	0
SKS prerequisites	0
Course prerequisites	Pass PBO, Pass AP
Course Start	08.30
Course Finish	10.30
Exam Schedule	Day 3, 10.30
Specialization Course	No

4.2 Crossover and Mutation Design

In this study, crossover design was made with the rule that a slot of a particular course can only be subjected to a cross for all of the time slots of the course, and crossed to all of the time slots of another course in the other chromosome. After that the crossover results need to be checked whether valid or did not. If invalid, then the result was not used in subsequent process. The example could be seen in Figure 5.

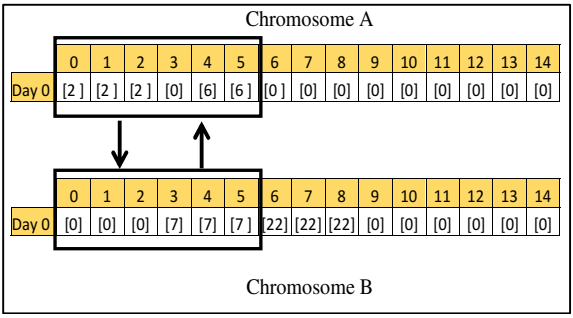


Fig. 5. The example of crossover

Whilst the mutation is done by randomly selecting two courses that have the same length of time in the one of crossover result chromosome, the offspring, and then swap their positions.

4.3 Fitness Cost

To determine the quality of the chromosome that is created, fitness cost calculation is required. Here are the fitness cost calculation formulas of a chromosome. We adopt the fitness cost calculation formulas from the similar formulas that was used by researchers in the previous work [7].

- Fitness Cost based on Credit Point Units/SKS (FCSKS)

$$FCSKS = \text{totalCreditPointUnit} * 30 \tag{1}$$

- Fitness Cost based on the semester to be taken (FCSM)

$$FCSM = \sum_{i=1}^{m} courseSemester[i] \tag{2}$$

Where:
If the course is mandatory for next semester:
courseSemester = (courseSemester [i] - semesters) * (-5)
If the course is mandatory for the last semester:
courseSemester = (semesters - courseSemester [i]) * (200)
If the course is mandatory for the current:
courseSemester = 100
m = number of existing courses

- Fitness Cost based on curriculum (FCKUR)

$$FCKUR = \sum_{i=1}^{m} courseCurriculum[i] \tag{3}$$

Where:
If the elective courses taken curriculum matches the user specialization:
courseCurriculum = 175
If the elective courses taken curriculum does not match the user specialization:
courseCurriculum = -75
m = number of courses in the chromosome

- Fitness Cost based on the interval between courses (FCJMK):

$$FCJMK = \sum_{i=1}^{n} dayInterval[i] \tag{4}$$

Where:
dayInterval = 300 - (interval between courses * 90)
n = number of day

- Fitness Cost based on courses to be taken (FCMKI)

$$FCMKI = \sum_{i=1}^{n} \text{ValueDesireCourse}[i] \tag{5}$$

Where:

n = number of existing courses on chromosome
If the user wishes 'Want to Take':
ValueDesireCourse = 100
If the user wishes 'Extremely Want to Take':
ValueDesireCourse = 1000
If the user wishes 'Less Want to Take':
ValueDesireCourse = -200

- Fitness Cost based on the time desired by the user (FCWU)

$$FCWU = \sum_{i=1}^{n} courseDay\text{Desired}[i] \quad (6)$$

$$courseDay\text{Desired} = \sum_{j=1}^{k} timeValue[j] \quad (7)$$

Where:
n = number of day
k = 3, on the morning: k = 1, on the afternoon: k = 2, on the night: k = 3
If in accordance with the expected time:
timeValue = number of hours filled * 200
If not in accordance with the expected time:
timeValue = number of hours filled * -200

- Fitness Cost based on the maximum number of courses per day that is desired by the user (FCMMK)
  — If user desirability ≥ number of courses,

$$FCMMK = \sum_{i=1}^{n} \frac{numberOfCourses\,[i]}{userDesirability[i]} * 300 \quad (8)$$

  — If user desirability ≥ number of courses

$$FCMMK = \sum_{i=1}^{n} \text{userDesirability[i]} - numberOfCourses[\text{i}] * 500 \quad (9)$$

Where:
n = number of day

- Furthermore, the formula to calculate overall fitness cost of a chromosomme is as follows:

$$\text{Fitness Cost} = FCSKS * w1 + FCSM * w2 + FCKUR * w3 + FCJMK * w4 + FCMKI * w5 + FCWU * w6 + FCMMK * w7 \quad (10)$$

Where:
w1 to w7 = weight of each fitness cost, value: 0...1.

5 Implementation and Experiment Results

System workflow could be seen in Figure 6.

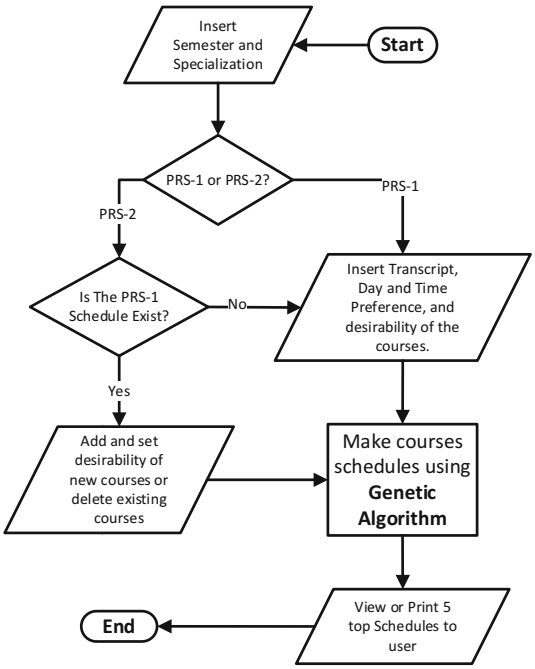


Fig. 6. System Workflow

System starts with a login form to identify the user of the application. Furthermore, the user can enter semester to be processed and areas of specialization are taken, as shown in Figure 7.

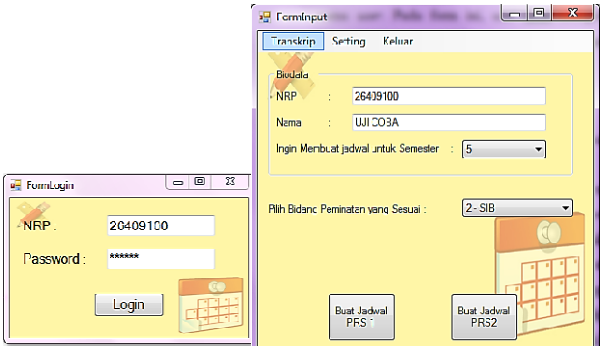


Fig. 7. Form Login and form setting semester and specialization

Later the user can choose to process the PRS-1 or PRS-2. Next will be displayed to the user some forms in sequence, as shown in Figure 8 to 10. Here the user can enter the latest transcript, day and time he desire and the desirability settings for all courses that could be taken for the next semester.

Fig. 8. Form to insert the latest transcript

Fig. 9. Form to choose preferences days and hours of courses, and the number of courses/day

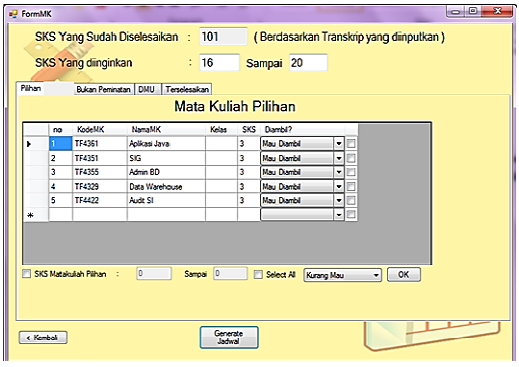


Fig. 10. Form to determine the desirability of all the courses that can be taken

Hereinafter the user can press a button to generate schedules using Genetic Algorithm. Once the process is complete then to the user would be presented 5 best schedules. These 5 best schedules are taken from 5 chromosomes with the highest fitness costs on the last population. The scheduling results form and also the look to be printed (HTML files) could be seen in Figure 11. In addition, before starting the process of scheduling, user can change the settings of variables required by the GA. Form to change the settings can be seen in Figure 12.

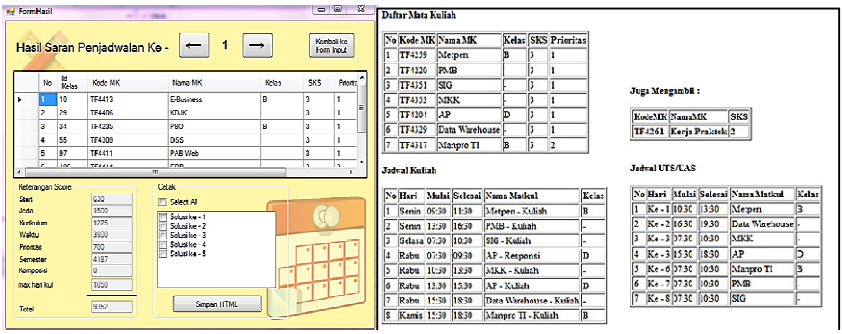


Fig. 11. Form to view 5 best the results of automatic courses scheduling and HTML files to print the schedule

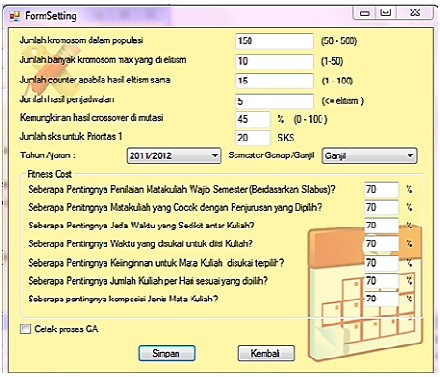


Fig. 12. Form setting for Genetic Algorithm process

To test the results of the system, especially the results of the courses schedule that is generated automatically, we use several kinds of tests. The first test is testing the effect of the maximum population against the processing time, and the best fitness cost value. We used the mutation probability = 0.3. The test results could be seen in Table 3.

Table 3. Tests on the maximum population size

Maximum Population	Running Times (millisecond)		Average of Fitness Costs
	Generate Initial Chromosomes	Genetic Algorithm Processes	
50	5962.8	3266.6	4302.4
150	15070.8	17938.4	4782.4
300	23283.2	32652.8	5297.4

From the test results it can be seen that the greater the maximum limit of the population, the longer the process takes time, yet the average value of the best fitness costs also increased. This is because more and more of the population there, will be the more diverse opportunities to generate chromosomes, so that eventually the fitness cost values of the chromosomes will increase.

The second test is the test of the effects of increased mutation probability values allowed on offspring chromosomes. For this test, the maximum population size be set at 150 chromosome. The test results can be seen in Table 4.

Table 4. Tests on the mutation probability values

Mutation Probability	Running Times (millisecond)		Average of Fitness Costs
	Generate Initial Chromosomes	Genetic Algorithm Processes	
0.1	11614	11935	4532.8
0.4	14621.4	12650	4712
0.7	15942.2	19960.6	5176

From the test results it can be seen that the greater the probability of mutation, the longer the time required for the process, but the average fitness value of a chromosome best cost also increased. This is because the higher the chance of mutation, the more diverse the resulting chromosome will be are greater, so that eventually the fitness cost value of the best chromosome will increase.

In addition we also did some other tests, such as testing the maximum number of chromosomes are copied by elitism method from the old population to the new population. We also tested the maximum number of 'elitism counter', this counter to count how many the sequences of populations will produce the same best fitness cost values, before the program stop the process. Of all the tests performed at last we obtained the best values of each setting. These values are used as the default values setting of the system. The amount of the default settings are: Maximum number of chromosome in the population = 150; Maximum number of chromosome copied to the new population by elitism procedure = 15; Maximum Elitism Iteration = 15; Mutation probability = 0.7. The testing results using default settings can be seen in Table 5.

Table 5. Testing results using default setting

Number of Tests	Running Times (millisecond)		Scores
	Generate Initial Chromosomes	Genetic Algorithm Processes	
1	23523	34653	5317
2	17116	18932	5317
3	17177	34653	5317
4	18592	32155	5269
5	33025	34090	5317
Average:	21886.6	30896.6	5307.4

Of testing by using the default settings could be seen that the GA produces the best fitness cost value greater compared to other settings that have been tested. Moreover, it can be seen that in general the system will be convergence on a particular chromosome.

The final testing is testing in the form of questionnaires to potential users, in example students majoring in Informatics department of Petra Christian University. The results summary of these questionnaires are as follows: As many as 60% of respondents gave scores of 4 and 40% gave scores of 5, of the scale from 1 (worst) to 5 (best), for ease of use of this application. From these results it can be concluded that the user interface of the application is easy to use. Meanwhile, for the question about the quality of the generated schedules, 70% of respondents gave scores of 4 and 30% gave scores of 5, of scale from 1 (worst) to 5 (best). Therefore, it can be concluded that in general the respondents felt be helped by the application that could generate courses schedule automatically.

6 Conclusion

From the tests it can be concluded that the automatic scheduling system is well made. Processing speed is also good, averaging less than 1 minute. From the test results on the potential users can be concluded that the interface has a good design and user friendly. In addition, the automatically generated class schedules are also correct and in accordance with the expectations of users.

References

1. Mitchell, M.: An Introduction to Genetic Algorithms. Massachusetts Institute of Technology, United States of America (1996)
2. Negnevitsky, M.: Artificial Intelligence: A Guide to Intelligent Systems, 2nd edn. Pearson Education Limited (2005)

3. Goldberg, D.E.: Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading (1989)
4. Davis, L.: Handbook on Genetic Algorithms. Van Nostrand Reinhold, New York (1991)
5. Haupt, R.L., Haupt, S.E.: Practical Genetic Algorithms, 2nd edn. John Wiley & Sons, Inc., United States of America (2004)
6. Holland, J.H.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975)
7. Budhi, G.S., Handojo, A., Soloment, B.: The Use of compact-Genetic Algorithm (cGA) for Petra Christian University Classroom Scheduling Optimization. In: Proc. National Seminar on Information Technology (2009)

Optimization of Auto Equip Function in Role-Playing Game Based on Standard Deviation of Character's Stats Using Genetic Algorithm

Kristo Radion Purba

Informatics Engineering - Petra Christian University
Surabaya, Indonesia
kristo@petra.ac.id

Abstract. Genetic algorithm is a well-known optimization solution for an unknown, complex case that cannot be solved using conventional methods.

In Role-Playing Games (RPG), usually the main features are character's stats and equip items. Character has stats, namely strength, defense, speed, agility, life. Also, equip items that can boost character's stats. These items retrieved randomly when an enemy dead.

A problem arise when the player have so many items that we cannot choose the best. Latest items doesn't always mean best, because usually in RPGs, items don't always boost all stats equally, but often it reduces certain stat while increasing the other.

Based on this, a function is built in this research, to auto equip all items, based on the standard deviation of character's stats after equipping. The genetic algorithm will evaluate the best combination of gloves, armors and shoes. This algorithm involves the process of evaluating initial population (items combination), selection, crossover, mutation, elitism, creating new population. The algorithm stops when the best fitness is getting stable in successive 3 generations.

After the auto equip process, the character is getting significantly stronger compared to using default equip items, measured by the remaining life after fighting with several enemies.

Keywords: Genetic algorithm, RPG game, optimization, equip item, artificial intelligence, standard deviation.

1 Introduction

Computer games, or better known as game is an application that uses electronic media, is a form of multimedia entertainment that is made interesting. Developing realistic characters and intelligent enemy is the main task in the modern game design, in particular, real-time strategy game (RTS), first-person shooter (FPS), and role-playing games. Non-player characters (NPC) must be able to provide enough challenge to attract players to defeat opponents (Lin, 2011:325), therefore the NPC should be given an intelligence. Artificial intelligence is the ability of computer to solve a problem that requires ingenuity if done by humans (Johnson, 2001).

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 64–75, 2015.
DOI: 10.1007/978-3-662-46742-8_6

The game designed in this research will implement a genetic algorithm (GA) to optimize equipment used by the enemy and player. GA is a type of artificial intelligence. Each enemy in this game has 3 types of equipment, namely gloves (glove), body protection (armor) and footwear (shoes). The GA will optimize equipment used by the enemy, so that the status of enemy forces evenly.

Besides the optimization done for the enemy, there is also auto equip function that works for the player character. The auto equip uses GA to find the best combination of equipment so that the standard deviation of character stats can become as small as possible. The use of standard deviation, is because of having to high in certain stats while very low in other stats often feels not balanced.

The use of auto equip function for player character will make players easier to play, while expert players still able to equip manually. The optimization done for the enemies are also contributing to the game's complexity and variation.

2 Literature Review

This research was conducted with reference to the studies that have been done before. To give some knowledge in general, concepts about role-playing game (RPG) and genetic algorithm will be explained in this chapter.

2.1 Role-Playing Game

A role-playing game (RPG) is a genre of video game where the gamer controls a fictional character (or characters) that undertakes a quest in an imaginary world. (Janssen, n.d.). Defining RPGs is very challeng-ing due to the range of hybrid genres that have RPG elements. Traditional role-playing video games shared three basic elements:

- Levels or character statistics that could be im-proved over the course of the game
- A menu-based combat system
- A central quest that runs throughout the game as a storyline

Modern and hybrid RPGs do not necessarily have all of these elements, but usually feature one or two in combination with elements from another genre.

2.2 Genetic Algorithm

In the computer science field of artificial intelligence, a genetic algorithm (GA) is a search heuristic that mimics the process of natural selection. This heuristic (also sometimes called a meta-heuristic) is routinely used to generate useful solutions to optimization and search problems. (Mitchell, 1996:2) Genetic algorithms belong to the larger class of evolutionary algorithms (EA). The mechanism of the EA is inspired by the evolutionary systems in biology, such as reproduction, mutation, recombination and selection. EA is divided into several techniques, which is genetic algorithm, genetic programming, evolutionary programming, etc. (Ashlock, 2004:9).

Genetic algorithms are something worth trying when everything else as failed or when we know absolutely nothing of the search space (Sivanandam, 2008: 22). In genetic algorithms, the term chromosome typically refers to a candidate solution to a problem, often encoded as a bit string. The "genes" are either single bits or short blocks of adjacent bits that encode a particular element of the candidate solution.
The general structure of a GA in computer science can be defined by the steps as follows (Basuki, 2003: 3):

- Generating the initial population, the initial population is generated randomly to obtain the initial solution.
- Selection of the individual, this process will evaluate each population to calculate the fitness value of each chromosome and evaluate it until stop-ping criterion is met. If the stop criterion is not met then the new generation will be formed again by repeating step 2.
- Reproduction and form a new generation, in the form used three operators mentioned above, namely reproduction operator/selection, crossover and mutation. This process is performed repeatedly to obtain a sufficient number of chromosomes to form a new generation in which this new generation represents the new solution.
- Crossover, used to cross the genes between couple, using certain probability. Crossover probability is a process to indicate a ratio of how many couples will be picked of mating. (Vivek and Narayanan, 2013:233)
- Mutation. In a binary representation, a mutation consists of flipping bits with a particular probability (Vivek and Narayanan, 2013:230)
- The process of elitism
  This process ensures that the generation of a population is not declining quality from generation to generation. This process will replace the worst individuals of the new population with the best individual in the population prior to reproduction.
- Stop at a certain generation.
  Stop after a few consecutive generations obtained the highest fitness value is not changed. Stop when the next generation n is not obtained higher fitness values.

2.3 Standard Deviation

The variance of a set of data from the mean that is, how far the observations deviate from the mean. This deviation can be both positive and negative, so we need to square these values to ensure positive and negative values do not simply cancel each other out when we add up all the deviations. (Math Centre, 2003:1). Standard deviation is the root of variance. The standard deviation, will be measured in the same units as the original data. The equation for standard deviation is given at equation (1).

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2} \quad (1)$$

3 Design Overview

In this chapter, will be explained about the game system is designed, and also how the application of the algorithm.

3.1 Character Stats

In the game that is designed in this research, the stats (statistics) of the player's character or enemies' characters that will be used are:

- Attack / **atk**
  Attack stat will reduce enemy's life when the attacker attacks its enemy. The maximum value is 99.
- Defense / **def**
  Is the percentage rate that will reduce the amount of incoming attack damage when enemy attacks. The maximum value is 99.
- Life / **life**
  Is the life value. If character run out of life, it will die. The maximum value is 999.
- Speed / **spd**
  Is the number of pixels / frame that represents the movement speed of the character. The maximum value is 99.
- Agility / **agi**
  A percentage of chance a character can dodge incoming enemy attacks. The maximum value is 99.
- Ammo
  The number of ammunition owned by enemy archers. The maximum value is 49.
- Experience
  Owned by player character only. The player character will get some number of experience when killing an enemy. The higher the experience, the higher the level of the player.

3.2 Enemy Types

In this game, there are 3 types of enemies, namely:

- Assaulter
  Has an aggressive nature. They will go forward as a front-liner enemy. Brings a sword weapon to attack from melee range. This enemy has the characteristics of strong attacks and having a medium amount of life.
- Archers
  Have an aggressive nature. Shoots arrows from a distant, hidden place. This enemy has the characteristics of strong attacks, high-speed movement, but a little life.
- Boss
  Has a very aggressive nature. Boss has a strong attack characteristics, having a large amount of life, but the slow movement speed.

Here is a table of the enemy types in the game, in Table 1.

Table 1. Enemy Types

Enemy Type	Unit Name	Figure
Assaulter	Knight	
Assaulter	Paladin	
Archer	Crossbowman	
Archer	Artemis	
Boss (Mission 1)	High-Priest	

At the beginning of the every mission, the program will generate enemies with random type in a predetermined position in the map. Every enemy will be given control to be able to move. After an enemy died, in a delay of 30 seconds, it will be resurrected with in the same position, with random type.

3.3 Character Equipment

Each character in this game will have equipments, which consists of three categories, namely gloves, armor, shoes. Whenever using a certain equipment, the stats the character will increase / decrease according to the equipment used. Here is a table of existing equipment in this game, in Table 2:

Table 2. Equipments

Name	**atk**	**def**	**spd**	**agi**
Iron Glove	+10	+10	0	0
Steel Glove	+15	+10	0	0
Dark Glove	+30	-10	0	0
Demon Glove	+40	-15	0	0

Table 2. (*Continued*)

White Glove	-10	+30	0	0
Holy Glove	-15	+40	0	0
Iron Armor	0	+25	-5	0
Steel Armor	0	+35	-10	0
Dark Armor	0	-10	+30	0
Demon Armor	0	-15	+40	0
White Armor	0	+30	-10	0
Holy Armor	0	+40	-15	0
Iron Shoes	0	0	+5	+10
Steel Shoes	0	0	+5	+15
Dark Shoes	0	0	+30	-10

4 Genetic Algorithm

In this chapter, will be explained the proposed genetic algorithm process to optimize the set of equipments that will be used in each characters.

4.1 Description

Here is a description of a genetic algorithm designed in this game:

- The player character and enemy has complete equipment set (including glove, armor, shoes). The genetic algorithms will perform the auto-equip with the following provisions:
  - For the player character: This equipment can be set manually by player, or player can use the "automatic equip" function, where the program will find the most optimal combination of equipment (with GA) for the player.
  - For the enemy: This equipment is set on each enemy when it first appeared to the map, optimally through the process of genetic algorithm.
- Equipment owned by the player character / enemy marked by a 3 digit of genes, ie genes that represent G (glove), A (armor), S (shoes). These gene numbers are numeric integer between 0-5 because there are 6 kinds of equipment for each type. The gene structure can be seen in figure 1 :

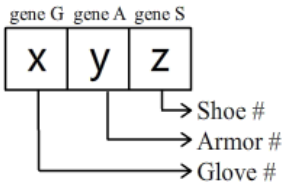


Fig. 1. Gene Structure

For example, the gene is written 213, then the equipment used by the character :

- Glove # 2, namely: Dark gloves
- Armor # 1, namely: Steel Armor
- Shoes # 3, namely: Demon shoes

4.2 Fitness Function

The fitness function is used to determine whether a gene is good or not, calculated using the standard deviation function, as explained in chapter 2.3, the implementation is shown in equation (2).

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{2}$$

where:

N = 4

X_1 = (**atk**) + (additional **atk** from equip item)

X_2 = (**def**) + (additional **def** from equip item)

X_3 = (**spd**) + (additional **spd** from equip item)

X_4 = (**agi**) + (additional **agi** from equip item)

$\bar{x}$ = Mean of x_0, x_1, x_2, x_3

The fitness value is the inverse of the standard deviation values. If the deviation is greater, then the fitness is getting smaller, and vice versa. Fitness value can be formulated based on the standard deviation, shown in equation (3).

$$f(x) = \frac{100}{\sigma+0.1} \tag{3}$$

With a standard deviation formula used in the calculation of fitness, best fitness is the minimum standard deviation. That is, it is expected that each of the character have a set of equipment that can create equity in character's stats, that is not too high at certain stat, but too low at another stat.

4.3 Genetic Algorithm Process

Suppose an assaulter has the stat: **atk** 12, **def** 15, **spd** 10, **agi** 7. Then, using a genetic algorithm, the program will find the best set of equipment that should be used by the character. This process is also known as auto equip. The cycle of the GA can be explained below.

1. Generate Random Individuals
   There are 10 individuals that will be generated, with random genes.
2. Calculate Fitness
   For example, for an individual with gene 445, it means that it equips:

- Glove # 4, namely: White gloves -> -10 **atk**, +30 **def**
- Armor # 4, namely: White armor -> +30 **def**, -10 **spd**
- Shoes # 5, namely: Holy shoes -> -15 **spd**, +40 **agi**

With the base stat of the character is **atk** 12, **def** 15, **spd** 10, **agi** 7, the fitness value can be calculated:

$$\bar{x} = \frac{(12\text{-}10) + (15\text{+}30\text{+}30) + (10\text{-}10\text{-}15) + (7\text{+}40)}{4} = 27.7$$

$$\sigma = \sqrt{\frac{1}{4}((12\text{-}10\text{-}27.7)^2 + (15\text{+}30\text{+}30\text{-}27.7)^2 + (10\text{-}10\text{-}15\text{-}27.7)^2 + (7\text{+}40\text{-}27.7)^2)} = 35.7$$

$$f(x) = \frac{100}{35.7 + 0.1} = 2.79$$

3. Calculate Probability
   The greater fitness value, the greater the probability of being selected in the process of roulette. Here is the formula for the probability, shown in equation (4).

$$prob_i = \frac{f(x)_i}{\sum_{i=1}^{n} f(x)_i} \tag{4}$$

   For example, here is the chosen couples that will be copulated:
   120 with 120, 120 with 212, 212 with 151, 151 with 212, 101 with 101

4. N-Point Crossover
   Crossover is done with chance of 70%. The higher the chance, the higher likeness of couple after copulation.
5. 1-Point Mutation
   After the crossover takes place, the result of crossover gene will be incorporated into the 1-point mutation, with chance of 15%. The higher the chance, the higher randomness of genes after mutation.
6. The Process Of Elitism
   The process of elitism is the process of replacing the worst gene of the current generation with the best gene from the previous generation, so the gene with best fitness can be maintained.
7. Next Iteration
   The process of genetic algorithm aims to find the most optimal set of equipment for a character. The GA will look for the higher fitness as possible. In the process of this GA iteration, the algorithm will stop when:
   - No improvement of best fitness value for 3 successive generations, or:
   - The number of generations has been more than 10. Limitation is intended that the game is not too heavy during the GA execution.

In Table 3 will be shown individual fitness changes from generation to generation until the genetic algorithm stops.

Table 3. Complete Process Example of Genetic Algorithm

Generation #					
1	2	3	4	5	6
445	101	220	125	121	121
423	120	125	121	125	121
120	220	251	125	125	321
212	012	125	125	120	121
411	100	125	122	121	101
151	125	100	125	155	121
212	252	100	125	121	121
101	111	122	105	121	121
502	112	101	125	121	101
553	251	122	225	123	421

In Table 3 shown that the genetic algorithm stops at the 7th generation, because of the fitness results have not improved during 3 successive generations. Therefore, can be concluded that the 024 is the best genes.

Conclusion:
For the character who has base stats of **atk** 12, **def** 15, **spd** 10, **agi** 7, the best equipment for it, is a representation of genes 121, namely:
- Gloves # 1, namely: Steel gloves
- Armor # 2, namely: Dark Armor
- Shoes # 1, namely: Steel shoes

5 Genetic Algorithm Results

After the genetic algorithm execution for several cases, the results will be documented in this chapter, divided by 2 categories, for the player character, and the enemy character.

5.1 Player Character

For the player character, the genetic algorithm is used for auto-equip menu, ie to automatically set the best equipment for player. In this section, the auto equip will be tested for 6 cases, as shown in Table 4. The best equipment is the set of equipment that produces good equalization status.

Table 4. GA result for player character

Base stats (with default equipment*)	Auto equip result	Stats after **auto equip**
attack = 60 **defense** = 60 **speed** = 25 **agility** = 25 (SD = 17.5)	- Iron gloves - Dark armor - Steel shoes	**attack** = 60 **defense** = 25 **speed** = 60 **agility** = 30 (SD = 16.3)
attack = 60 **defense** = 60 **speed** = 100 **agility** = 20 (SD = 28.3)	- Steel gloves - Steel armor - Steel shoes	**attack** = 65 **defense** = 70 **speed** = 95 **agility** = 25 (SD = 25.1)
attack = 35 **defense** = 60 **speed** = 25 **agility** = 90 (SD = 25.1)	- Dark gloves - Steel armor - Demon shoes	**attack** = 55 **defense** = 50 **speed** = 55 **agility** = 65 (SD = 5.4)
attack = 100 **defense** = 60 **speed** = 25 **agility** = 25 (SD = 30.9)	- Iron gloves - Dark armor - Steel shoes	**attack** = 100 **defense** = 25 **speed** = 60 **agility** = 30 (SD = 29.9)
attack = 35 **defense** = 100 **speed** = 25 **agility** = 25 (SD = 31.3)	- Dark gloves - Demon armor - Holy shoes	**attack** = 55 **defense** = 40 **speed** = 50 **agility** = 55 (SD = 6.1)
attack = 40 **defense** = 60 **speed** = 80 **agility** = 80 (SD = 16.6)	- Dark gloves - Holy armor - Iron shoes	**attack** = 60 **defense** = 55 **speed** = 70 **agility** = 80 (SD = 9.6)

* Default equipment is Glove #1, Armor #1, and Shoes #1

5.2 Enemy Character

The auto equip for enemy character is done at the first time it appeared to the map. The results of genetic algorithm for enemy is shown at table 5.

Table 5. GA result for enemy character

Base stats (with default equipment*)	Auto equip result	Stats after **auto equip**
attack = 14 **defense** = 53 **speed** = 11 **agility** = 16 (SD = 17.2)	- *Steel Glove* - *Dark Armor* - *White Shoes*	**attack** = 26 **defense** = 18 **speed** = 31 **agility** = 36 (SD = 6.6)
attack = 15 **defense** = 59 **speed** = 12 **agility** = 16 (SD = 19.4)	- *Steel Glove* - *Demon Armor* - *White Shoes*	**attack** = 30 **defense** = 24 **speed** = 27 **agility** = 46 (SD = 8.5)
attack = 17 **defense** = 44 **speed** = 14 **agility** = 21 (SD = 11.8)	- *White Glove* - *Dark Armor* - *White Shoes*	**attack** = 12 **defense** = 29 **speed** = 34 **agility** = 41 (SD = 10.7)
attack = 19 **defense** = 44 **speed** = 14 **agility** = 28 (SD = 11.4)	- *Iron Glove* - *Iron Armor* - *Steel Shoes*	**attack** = 38 **defense** = 44 **speed** = 14 **agility** = 33 (SD = 11.2)
attack = 18 **defense** = 55 **speed** = 16 **agility** = 20 (SD = 16)	- *Iron Glove* - *Dark Armor* - *White Shoes*	**attack** = 35 **defense** = 20 **speed** = 36 **agility** = 40 (SD = 7.6)

* Default equipment is Glove #1, Armor #1, and Shoes #1

6 Conclusion and Suggestion

6.1 Conclusion

From the results shown in chapter 5, can be seen that genetic algorithm can be applied to optimize the equipment used by the player and the enemy, based on the stats of each player and the enemy. Each player and the enemy has the status of attack, defense, speed, agility. Genetic algorithm will find a combination of equipment (glove, armor, shoes) the best that can make stats that doesn't deviate too much. The fitness value used in the genetic algorithm is the invers of standard deviation formula. The higher the standard deviation, the lower the fitness, and vice versa.

6.2 Suggestion

Here are some suggestions to develop this research to be better:

- Add more enemy types and equipment types in the game so that searching space for genetic algorithm is more extensive, so the genetic algorithm can be more useful.
- Also use the sum of stats for the fitness function, not just standard deviation, because players usually also consider the stats number when equipping character.

References

1. Ashlock, D.: Evolutionary Computation for Modeling and Optimiza-tion, p. 9. University of Venice, Italia (2004)
2. Basuki, A.: Algoritma Genetika (2003), http://www.pens.ac.id/~basuki/lecture/AlgoritmaGenetika.pdf (retrieved March 25, 2013)
3. Janssen, C. (n.d.).: Role-Playing Game, http://www.techope-dia.com/definition/27052/role-playing-game-rpg (retrieved December 4, 2014)
4. Johnson, D., Wiles, J.: Computer Games with Intelligence. Australian Journal of Intelligent Information Processing Systems, 61 (2001)
5. Lin, C.-S., Ting, C.-K.: Emergent Tactical Formation Using Genetic Algorithm in Real-Time Strategy Games. In: Conference on Technologies and Applications of Artificial Intelligence, p. 325 (2011)
6. Math Centre. Variance and standard deviation (ungrouped data) (2003)
7. Mitchell, M.: An Introduction to Genetic Algorithm. MIT Press, London (1996)
8. Sivanandam, S.N., Deepa, S.N.: Introduction to Genetic Algorithms, p. 22. Springer, New York (2008)
9. Vivek, G., Narayanan, R.C.: Minimization of Test Sequence Length for Structural Coverage Using VLR. IJRIT International Journal of Research in Information Technology 1(5), 220–227 (2013)

The Design of Net Energy Balance Optimization Model for Crude Palm Oil Production

Jaizuluddin Mahmud[1,2], Marimin[2,3], Erliza Hambali[2,3], Yandra Arkeman[2,3], and Agus R. Hoetman[4]

[1] Research Centre for Technology Innovation Policy
Agency for Assessment and Application of Technology (BPPT), Jakarta, Indonesia
jaizoel@yahoo.com
[2] Department of Agroindustrial Technology, Faculty of Agricultural Technology,
Bogor Agricultural University (IPB), Bogor, Indonesia
[3] Surfactant and Bioenergy Research Center,
Bogor Agricultural University (IPB), Bogor, Indonesia
[4] Deputy Minister for Science and Technology Network
State Ministry of Research and Technology, Jakarta, Indonesia

Abstract. Net energy balance (NEB) is the second important indicator following green house gases in developing a sustainable biodiesel industry. The extent of the production chain and various ways to reduce the use of fossil fuels, increase the complexity of finding an optimal NEB value of the industry. The main objective of this study was to design an NEB optimization model, which was supported by genetic algorithm (GA). The model was applied in a crude palm oil (CPO) industry that produces raw material for biodiesel which is located in North Sumatra province. The model was solved by using an optimization computer software package. The results showed that the NEB value was better than the previous one. The model was also able to provide biomass usage composition to achieve the optimal NEB value, and the unit processes that need to be improved.

Keywords: biodiesel, CPO, net energy balance, optimization model, genetic algorithm.

1 Introduction

Recently, biodiesel have become international issues in addressing energy security problems, decreasing the dependency on fossil fuels, and reducing greenhouse-gas (GHG) emissions [1,2]. In many countries, biodiesel has been used as an alternative energy to substitute diesel fuel. Furthermore, compared to diesel fuel, biodiesel has several advantages. Biodiesel produces no sulfur, no net carbon dioxide, less carbon monoxide, particulate matters, smoke and hydrocarbons emission and more oxygen. It can replace diesel fuel in many different applications, transportation and industry such as in boilers and internal combustion engines [3].

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 76–88, 2015.
DOI: 10.1007/978-3-662-46742-8_7

In order to support the reduction of dependence on fossil fuel, biodiesel industry should also carry out the reduction of the conventional energy use. But in fact, it is difficult for the industry to avoid the use of this fuel due to its dependence on fertilizers and methanol industry and transport [4]. However, the use of fossil fuel in the biodiesel industry should be kept to a minimum, at least the energy balance is maintained to not use fossil fuel in excess of biodiesel produced. The net energy balance (NEB) became the second important indicator following greenhouse-gas (GHG) in the analysis of the sustainability of a biodiesel industry [5,6,7].

All NEB studies on biodiesel industry showed a positive value [8],[17]. De Souza et al. [8] reported their study that from 6 studies on energy balance showed different energy ratios, i.e. 4.99, 4.68, 4:05, 4:58, 3:15, and 6.69. largest energy consumption is derived from palm oil mill, which averaged 50%, respectively following the refinery and biodiesel plant, plantation, transportation and energy inputs respectively 24%, 16%, and 10%. If produce CPO includes stages plantation, transportation and palm oil mill, then the total energy consumption to 76% of the total energy consumption of biodiesel industry.

Based on the description mentioned above, it is important to discuss the net energy balance in CPO production stage, especially Indonesia as well known in the world's largest palm oil producer. Some studies have suggested raising the NEB through improved energy use in plantations and CPO mill [9,10,11,12],[18]. Study by Kawahara et al. in Lampung - Indonesia [10] reported potential energy improvement of 3.2 MJ kg^{-1} through use of biomass waste from palm oil mill. Pleanjai and Gheewala [11] reported that by recovering biogas from anaerobic treatment of palm oil mill wastewater, the NER increasing from 3.58 to 3.70.

Almost all NEB studies on biodiesel industry showed a positive value, or in other words, the biodiesel industry uses fossil fuels less than biodiesel produced [8]. Some studies even suggest to increase the value of the NEB industry by the use of alternative fuels, such as solid waste biomass of palm oil mill [9,10], and biogas palm oil mill effluent [11,12].The main objective of this study was to design an NEB optimization model of CPO industry in producing raw material for biodiesel, which is supported by genetic algorithm (GA).

2 Energy Balance System

2.1 CPO Production Process

Description of the production process of CPO refers to life cycle assessment (LCA) which has been generated from several studies [13,14]. Based on the LCA, there are two stages of the production process, plantation and palm oil mill. Plantation produces fresh fruit bunches (Fresh Fruit Bunches / FFB) from the age of 2 or 3 years, and then continued to bear fruit until the age of 20-25 years. To growing well, the plant requires maintenance, especially the provision of inorganic fertilizer supplied from outside the plantation. The harvested ripe fruit, taken to storage warehouse to be transported to the palm oil mill.

To produce CPO, FFB is processed by sterilization, stripping, and extraction. During the process, the plant produces solid and liquid wastes. Solid waste includes Empty Fruit Bunches (EFB), fibers (mesocarp fiber, MF), palm kernel and shell (palm kernel shell, PKS). The kernel can be recycled into crude palm kernel oil (CPKO).

EFB, MF and PKS used as biomass are burned in a boiler to produce steam and electricity for the purposes of plant. EFB is generally brought back to the farm as fertilizer. Meanwhile, with a particular treatment plant effluent can be used as biogas and liquid fertilizer as a substitute for organic fertilizer on the farm.

2.2 Net Energy Balance

In this study, the energy balance system aims to maintain sustainability of CPO production process in terms of energy use, in other words that the potential energy possessed CPO must be greater than the energy used. Energy output is represented by the CPO and the kernel, while the input energy is divided into two groups, which are direct and indirect energy. Direct energy is energy that is burned or converted directly on the CPO mill, both at the stage of cultivation and processing. Direct energy can be a fossil fuel (gasoline, diesel oil), biofuels, biomass, biogas and electricity. They are used in all units of the production process in plant, transportation of FFB from

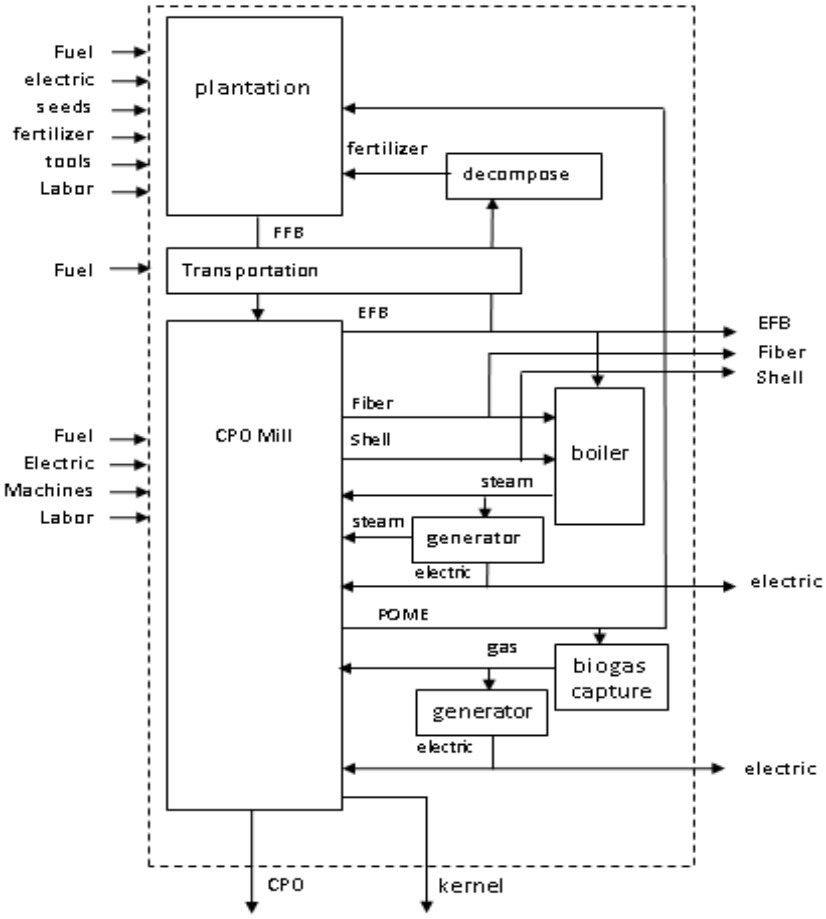


Fig. 1. Energy input output system net energy balance production of crude palm oil (CPO)

plantation to the mill, and EFB from the mill to the plantation. Indirect energy is energy that is attached to the material (embodied energy) because it requires energy to be produced and delivered to the plantation or plant. Indirect energy includes the provision of seeds, fertilizers, heavy equipment machinery. Energy inputs and outputs on the net energy balance system CPO production are shown in Figure 1.

3 System Modelling

3.1 NEB System Modeling

There are several values generated by calculation of net energy balance of CPO production, i.e. the value of net energy system (NES), net energy balance (NEB), and net energy ratio (NER). NES consider all types of energy input, while the NEB only fossil energy. NEB is divided into two phases, existing and optimal condition. All calculations refer to the CPO kg weight. NES, as defined in equation (1), is equal to energy output (E_{out}) minus energy input (E_{in}), measured in MJ/ kg CPO.

$$NES = E_{out} - E_{in} \tag{1}$$

Energy input is divided into 3 groups, i.e. the energy used in the plantations, mill, and transportation. Energy input in plantation consist of 3 variables, fuel, electricity and fertilizer production. Fertilizer can be in the form of inorganic and organic that are produced from EFB and POME. Energy input in the mill consists of 3 types, i.e. fossil fuel, steam, and electricity. Electrical consumption of mill is supplied from grid, co-cogeneration and biogas. There are two kinds of transportation, taking FFB from farm to mill, and EFM from mill to farm.

$$E_{out} = E^{CPO} + E^{kern} \tag{2}$$

$$E_{in} = E_{in}^{plant} + E_{in}^{mill} + E^{tr} \tag{3}$$

$$E_{in}^{plant} = E_{fuel}^{plant} + E_{el}^{plant} + E_{fert}$$

$$E_{fert} = \sum_{j=1}^{n} E_{inorg,j}^{fert} + \sum_{j=1}^{n} E_{efb,j}^{fert} + \sum_{j=1}^{n} E_{pome,j}^{fert}$$

$$E_{in}^{mill} = E_{fuel}^{mill} + E_{steam}^{mill} + E_{el}^{mill}$$

$$E_{el}^{mill} = E_{grid}^{mill} + E_{efb}^{mill} + E_{mf}^{mill} + E_{pks}^{mill} + E_{biogs}^{mill}$$

$$E^{tr} = E_{ffb}^{tr} + E_{efb}^{tr}$$

In existing state, NES, NEB and NER values obtained of equation (4) (5) and (6) respectively.

Net Energy of System

$$= (E^{CPO} + E^{kern}) + E_{fuel}^{plant} + E_{el}^{plant} - \left(\begin{array}{c} \sum_{j=1}^{n} E_{inorg,j}^{fert} + \sum_{j=1}^{n} E_{efb,j}^{fert} + \sum_{j=1}^{n} E_{pome,j}^{fert} + E_{fuel}^{mill} + E_{steam}^{mill} + E_{grid}^{mill} + E_{efb}^{mill} + E_{mf}^{mill} \\ + E_{pks}^{mill} + E_{biogs}^{mill} + E_{ffb}^{tr} + E_{efb}^{tr} \end{array} \right) \quad (4)$$

Net Energy Balance (existing)

$$NEB = (E^{CPO} + E^{kern}) - \left(E_{fuel}^{plant} + E_{el}^{plant} + \sum_{j=1}^{n} E_{inorg,j}^{fert} + E_{fuel}^{mill} + E_{grid}^{mill} \right) \quad (5)$$

Net Energy Ratio (existing)

$$NER = (E^{CPO} + E^{kern}) / \left(E_{fuel}^{plant} + E_{el}^{plant} + \sum_{j=1}^{n} E_{inorg,j}^{fert} + E_{fuel}^{mill} + E_{grid}^{mill} \right) \quad (6)$$

3.2 NEB Optimization Model

Objective Function. Optimization, represented by net energy ratio (NER), is calculated by maximizing the NER value, through the reduction of the use of fossil fuels in all of stage of CPO production. Objective function, which is show in equation (7), has 12 variables considered, i.e. CPO and palm kernel; fertilizer; the use of diesel oil in plantation, transportation and mill; electricity for operation of plantation; electricity in the mill, and the use of EFB MF PKS and POME. If the variables in the optimization condition are denoted by 'op', then the equation optimization NEB is as follows:

Maximize NEB =

$$\frac{(E^{CPO} + E^{kern})}{\left(\begin{array}{c} E_{fuel}^{plant} + E_{el}^{plant} + \sum_{j=1}^{n} (E_{fert} - E_{efb.op,j}^{fert} - E_{pome.op,j}^{fert}) \\ + \sum_{k}^{m} E_{k}^{tr} + E_{fuel}^{mill} + E_{el}^{mill} - \left(E_{efb.op}^{mill} + E_{mf.op}^{mill} + E_{pks.op}^{mill} + E_{biogs.op}^{mill} \right) \end{array} \right)} \quad (7)$$

Constraints. Fossil fuel reduction is carried out by substitution with using alternative energy. Diesel oil, electricity, and fuel for fertilizer production, are substituted by biodiesel and biomass. NER can also be increased by using liquid waste (POME) to substitute inorganic fertilizer. Optimization using biodiesel can be calculated directly, but need proper combination using biomass produced by mill, such as EFB, MF, PKS and POME.

$$E_{el}^{mill} \geq E_{efb.op}^{mill} + E_{mf.op}^{mill} + E_{pks.op}^{mill} + E_{biogs.op}^{mill}$$

$$Fert_j \geq Fert_{efb.op,j} + Fert_{pome.op,j}$$

$$EFB \geq \sum_{j=1}^{n} EFB_{fert.op,j} + EFB_{mill.op}$$

$$Fert_{efb.op,j} + Fert_{pome.op,j}, EFB_{mill.op}, MF_{mill.op}, PKS_{mill.op}, Biogs_{mill.op} \geq 0$$

4 Genetic Algorithm

Genetic algorithm is a search technique and optimization techniques which work by mimicking the process of evolution and changes in the genetic structure of living things. Genetic algorithm was first introduced in the 1970s by John Holland, and then developed by scientists in solving complex optimization problems. Currently, Genetic algorithm has been applied in various fields, including engineering, investment, robotics, and industrial management.

The first step to do is to make encoding a candidate solution into a form of representation of chromosomes. Genetic algorithm works by measuring how well a chromosome can solve a problem. measurements were performed by using a fitness function. Genetic Algorithms solve the optimization problems by manipulating the chromosomes *blindly* using *crossover* and *mutation* operators.

The steps to implement the genetic algorithm are as follows:

1) Generate initial population randomly
2) Evaluate each chromosome in this population using its fitness function
3) Select a set of chromosomes to be "parents" and then reproduce new chromosomes using crossover and mutation operators
4) Evaluate these new chromosomes using their fitness functions
5) If stopping criteria reached, then stop else go to step 3.

5 Case Study

The model was applied to a CPO company located in Sumatera – Indonesia. The private company has a production capacity of 3.78 t.cpo ha^{-1} and production capacity 25 t/h = 500 t/d. The factory received FFB at 277,002.13 MT, produced CPO of 19.85% and a co-product kernel 5:20%. The factory produced solid waste EFB, MF, and PKS with the percentage of 22.00, 14.44, and 6.46 percent of the weight of FFB. The allocation of material inputs and outputs in per metric ton CPO can be seen in Table 1. Energy coefficients used in this study are taken from some references (Table 2). The coefficient of CPO and palm kernel is assumed same as the value of biodiesel, which is equal to 39.60 MJ/kg.

Table 1. Energy input and output in CPO production

No.	Item	Real data per year		Per Metric Ton CPO per year	
		Value	Unit	Value	Unit
1	Input				
	FFB	277,002.13	MT	4.70	kg/MT CPO
	Electricity (from grid)	95,799.00	kWh	1.62	kWh/MT CPO
	Diesel used for Diesel Generator	56,366.00	L	0.96	L/MT CPO
	Diesel used for transport EFB	27,871.00	L	0.47	L/MT CPO
	Diesel for transport FFB	773.82	L	0.12	L/MT CPO
	NPK - Super K	1,574.466	kg	1257.12	kg/MT CPO
	Dolomite	133,116	kg	106.29	kg/MT CPO
	Kieserite	145.288	kg	116.00	kg/MT CPO
	Urea	331.805	kg	264.93	kg/MT CPO
	MOP	449.021	kg	358.52	kg/MT CPO
	RP	100.583	kg	80.31	kg/MT CPO
	Output				
2	Main product				
	CPO	58,991.36	MT	1,000.00	kg/MT CPO
	By product				
	EFB	60,940.47	MT	1,033.04	kg/MT CPO
	MF	51.11	MT	0.87	kg/MT CPO
	PKS	9,203.08	MT	156.01	kg/MT CPO
	Palm Kernel	14.404	MT	244.17	kg/MT CPO
	POME	160.661	MT	2,723.47	kg/MT CPO

Table 2. Energy coefficients

Item	Unit	Value
CPO	MJ/kg	39.60
Palm kernel	MJ/kg	39.60
PKS [8]	MJ/kg	11.91
MF [8]	MJ/kg	7.7
EFB [15]	MJ/kg	7.24
Diesel [16]	MJ/L	40.3
Fertilizer [10]		
NPK - Super K	MJ/kg	2.8
Dolomite	MJ/kg	0.5
Kieserite	MJ/kg	2
Urea	MJ/kg	33.2
MOP	MJ/kg	3.5
RP	MJ/kg	1.3
Electricity[11]	MJ/KWH	9.9
Biogas [8]	MJ/kg	1.8

Table 3 shows the structure of energy for the company in producing CPO. The company uses energy (fuel oil) 14.061 MJ / yr kg CPO to produce 49.269 MJ / yr kg CPO. In existing condition, NES, NEB, and NER is equal 33.34 MJ / yr kg CPO, 35.21 MJ / yr kg CPO, 3.504 respectively. The NES of system that considered of biomass as input energy, by the greatest energy use occurs in fertilization, that was 14.333 MJ/yr kg CPO (87.%), followed by electrical 1.88 MJ/yr kg CPO (11.82%), and transportation 0.046 MJ/yr kg CPO (0.29%) from all of input energy in the system.

Table 3. Energy in CPO production

Item	**MJ / yr kg CPO**	**Item**	**MJ / yr kg CPO**
Input		b. CPO Mill	
a. Plantation		Electricity (from grid)	0.017
NPK - Super K	3.520	Diesel used for Diesel Gen	0.039
Dolomite	0.053	Diesel used for transport EFB	0.019
Kieserite	0.232	(b) Sub-total	0.075
Urea	8.796	Total Input (a + b)	**14.061**
MOP	1,255	**Output**	
RP	0.104	CPO	39.600
Diesel used (for transport FFB)	0.027	Palm Kernel	9.669
(a) Sub-total	13.987	Total Output	**49.269**

6 Result and Discussion

6.1 Designing Genetic Algorithm

In designing a genetic algorithm, there are four things to consider, namely chromosome representation, fitness function, selection of chromosomes, and genetic algorithm operators [19].

Chromosome Representation. To apply the genetic algorithm, the first step should be done is make an encoding candidate solutions in the form of representation of chromosomes. Type of encoding used in this study is binary.

Fitness Function. Genetic algorithm works by measuring how well a chromosome can resolve the problem. Measurements were made using a fitness function, which is the objective function of the problem to be solved. The following is the fitness function (8) and the constraints of this study:

Maximize NER =

$$\frac{39.6(\mathrm{CPO}+\mathrm{PK})}{\left(\begin{array}{c}40.3\ \mathrm{Fuel}^{\mathrm{farm}}+10.47\ \mathrm{Elec}^{\mathrm{farm}}+\sum_{i=1}^{n} a_i(\mathrm{Fert}_{\mathrm{total.i}}-\mathrm{Fert}_{\mathrm{efb.i}}-\mathrm{Fert}_{\mathrm{pome.i}})+\\ \sum_{j}^{m}(40.3-0.7\ b_j)\ \mathrm{Fuel}_j^{\mathrm{tr}}+40.3\ \mathrm{Fuel}^{\mathrm{mill}}+\mathrm{E}_{\mathrm{elec.total}}^{\mathrm{mill}}-\\ 18.1\ \mathrm{EFB}-7.7\ \mathrm{MF}-11.91\ \mathrm{PKS}-1.8\ \mathrm{Biogs}\end{array}\right)} \quad (8)$$

Constraints:

$$\mathrm{CPOm} \geq \mathrm{CPO}$$

$$\mathrm{PKm} \geq \mathrm{PK}$$

$$\mathrm{PKm} = \mathrm{PKe}/\mathrm{CPOe}$$

$$\mathrm{EFBm} \geq \mathrm{EFB}$$

$$\mathrm{EFBm} = \mathrm{EFBe}/\mathrm{CPOe}$$

$$\mathrm{MFm} \geq \mathrm{MF}$$

$$\mathrm{MFm} = \mathrm{MFe}/\mathrm{CPOe}$$

$$\mathrm{PKSm} \geq \mathrm{PKS}$$

$$\mathrm{PKSm} = \mathrm{PKSe}/\mathrm{CPOe}$$

$$\mathrm{POMEm} \geq \mathrm{POME}$$

$$\mathrm{POMEm} = \mathrm{POMEe}/\mathrm{CPOe}$$

$$\mathrm{Fert}_{\mathrm{efb.i}} = b_{\mathrm{efb.i}}\mathrm{EFB}_{\mathrm{fert}}$$

$$\mathrm{Fuel}m \geq \mathrm{Fuel}$$

$$\mathrm{Fuel}e = \mathrm{Fuelm}$$

$$\mathrm{E}_{\mathrm{elec.total}}^{\mathrm{mill}} \geq 18.1\ \mathrm{EFB}_{\mathrm{el}} - 7.7\ \mathrm{MF} - 11.91\ \mathrm{PKS} - 1.8\ \mathrm{Biogs}$$

$$\mathrm{EFB}_{\mathrm{fert}} + \mathrm{EFB}_{\mathrm{el}} \leq \mathrm{EFB}$$

$$\mathrm{POME}_{\mathrm{fert}} + \mathrm{POME}_{\mathrm{el}} \leq \mathrm{POME}$$

$$\mathrm{Fert}_{\mathrm{total.i}} \geq \mathrm{Fert}_{\mathrm{efb.i}} + \mathrm{Fert}_{\mathrm{pome.i}}$$

$$\mathrm{CPO}, \mathrm{PK}, \mathrm{Fuel}, \mathrm{Fert}_{\mathrm{efb.j}}, \mathrm{Fert}_{\mathrm{pome.i}}, \mathrm{POME}_{\mathrm{fert}}, \mathrm{EFB}_{\mathrm{fert}}, \mathrm{EFB}_{\mathrm{el}}, \mathrm{MF}, \mathrm{PKS}, \mathrm{Biogs} \geq 0$$

Selection Chromosome. Selection is one of the operation to ensure that the number of representatives of a chromosome that is received in the next generation will depend on the fitness value compared to the value of the average fitness of the population. Chromosome which has a very good fitness value will have a greater opportunity for the parent and elected to remain in the next generation, while the poorer chromosome will be replaced by a new chromosome. There are two ways of selection are known in the GA, the fitness proportional selection and roulette wheel selection. The study used roulette wheel selection, where each chromosome in the population occupies a slot on

the roulette disc. The large size of the slot is equal to the ratio between the value of the fitness of a chromosome with a total value of fitness of all chromosomes.

GA Operator. The main objective to be achieved in the use of genetic algorithms is that a set of chromosomes (candidate solutions) are generated randomly in the initial population and multiply by itself to converge to provide a best fitness value (optimum value). Chromosomes are formed in this new generation of are called offspring. An offspring can be shaped by two the main processes, crossover and mutation. Crossover is the main operator in genetic algorithm. Opportunities crossover is the ratio between the numbers of chromosomes that is expected to experience a crossover in each generation with the total number of chromosomes in the population. Therefore, the crossover is the primary operator, crossover probability value is usually quite high (0.6 - 1).

Mutation is supporting operators in genetic algorithm that acts to change the structure of chromosomes spontaneously. This change led to the formation of a spontaneous mutant, which is a new chromosome is genetically different from the previous chromosome. Opportunities mutation is the ratio between the expected numbers of mutated genes in each generation with the total number of genes in a population. Therefore, mutations are supporting the operator, the probability of which is used usually quite low (0.001-0.2).

The values of genetic algorithm parameters used for this case are as follows:

1. Population size = 20
2. Crossover probability = 0.9
3. Mutation probability = 0.01
4. Maximum generation = 100

6.2 Implementation Model with GA

Besides the aim to search the optimal value of the NEB, the model is also expected to show the best value of any variable. Basically all of the variables in fitness function can be optimized to get the best NEB, but because of the difficulty of data then the model in this case to consider five variables, namely EFB and POME to replacing inorganic fertilizer, and EFB MF and PKS to substitute electrical from grid.

Implementation model with genetic algorithms was supported by MATLAB. The model produced the optimum value of the NEB by 36.258 MJ/kg CPO, increased 1.05 MJ/kg CPO from the previous NEB. NER optimization was 3.787, an improvement of 8.08% from the previous NER.

Increasing the value of the NEB obtained from variables of EFB and POME for fertilizer, EFB MF and PKS for electrical in the mill. The use mill waste to the reduction of fossil fuel is shown in Table 4 and Fig.2. Even without using of mutation operators, achieving the optimization value gained up to 51 iterations. This best optimization value is shown in Fig. 3.

Table 4. The use mill waste for optimization NEB

Variables	Value (kg/kg_CPO)
Fertilizer	
1. EFB	1.033
2. POME	2.724
Electrical	
3. EFB	0
4. MF	0.244
5. PKS	0

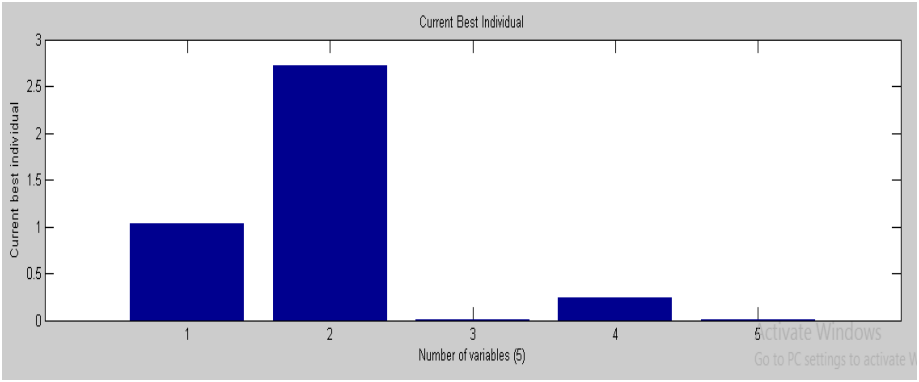


Fig. 2. The best value of each variable for NEB optimization

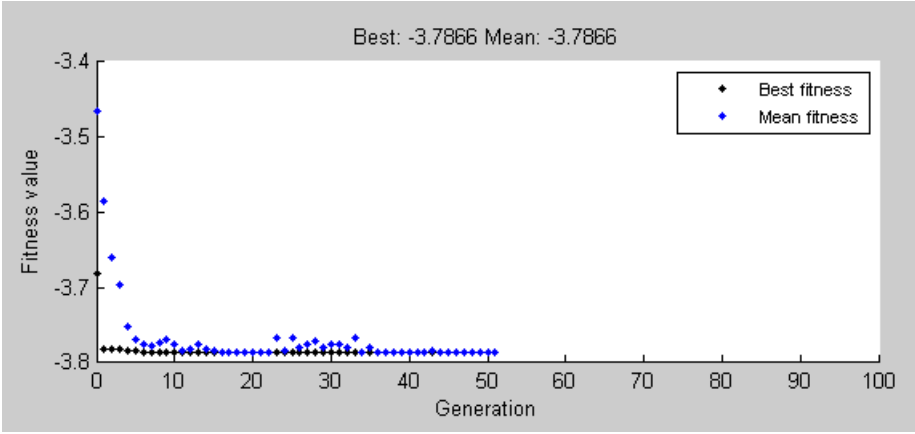


Fig. 3. The best fitness of NEB optimization

For further study, the developed model can be expanded for solving fuzzy values of the decision variables and constraint by using fuzzy approach such as suggested in [20].

7 Conclusion and Recommendation

This model was able to provide information about the optimal value of the NEB that can be achieved in the manufacturing process of CPO, and the sources of energy that can be optimized in order to minimize the use of fossil fuel in the company. Fertilizer and the use of electricity in the plant, were two most influential components in determining the performance of the net energy balance optimization in producing CPO. CPO mill waste has a significant role in improving the performance of the NEB optimization.

Model optimization of net energy balance needs to be continued at another stage in the production chain of the biodiesel industry.

References

1. Linares, P., Peŕez-Arriaga, I.J.: A Sustainable Framework for Biofuels in Europe. Energy Policy 52, 166–169 (2013)
2. Doku, A., Di Falco, S.: Biofuels in Developing Countries: Are comparative advantages enough? Energy Policy 44, 101–117 (2012)
3. Atabani, A.E., Silitonga, A.S., Badruddin, I.A., Mahlia, T.M.I., Masjuki, H.H., Mekhilef, S.: A Comprehensive Review on Biodiesel as an Alternative Energy Resource and Its Characteristics. Renewable and Sustainable Energy Reviews 16, 2070–2093 (2012)
4. Fore, S.R., Porter, P., Lazarus, W.: Net Energy Balance of Small-Scale on-Farm Biodiesel Production From Canola And Soybean. Biomass and Bioenergy 35, 2234–2244 (2011)
5. Pradhan, A., Shrestha, D.S., Gerpen, J.V., Duffiield, J.: The Energy Balance of Soybean Oil Biodiesel Production: A review of past studies. ASABE 5, 185–194 (2008)
6. Buchholz, T., Luzadis, V.A., Volk, T.A.: Sustainability Criteria for Bioenergy Systems: Results From an Expert Survey. Journal of Cleaner Production 17, S86–S98 (2009)
7. Markevicius, A., Katinas, V., Perednis, E., Tamasauskiene, M.: Trends and Sustainability Criteria of the Production and Use of Liquid Biofuels. Renewable and Sustainable Energy Reviews 14, 3226–3231 (2010)
8. de Souza, S.P., Pacca, S., Ávila, M.T., Borges, J.L.B.: Green House Gas Emissions and Energy Balance of Palm Oil Biofuel. Renewable Energy 35, 2552–2561 (2010)
9. Shuit, S.H., Tan, K.T., Lee, K.T., Kamaruddin, A.H.: Oil Palm Biomass as a Sustainable Energy Sources: A Malaysia Case Study. Energy 34, 1225–1235 (2010)
10. Kamahara, H., Hasanudin, U., Widiyanto, A., Tachibana, R., Atsuta, Y., Goto, N., Daimon, H., Fujie, K.: Improvement Potential for Net Energy Balance of Biodiesel Derived From Palm Oil: A Case Study from Indonesian Practice. Biomass and Bioenergy 34, 1818–1824 (2011)
11. Pleanjai, S., Gheewala, S.H.: Full Chain Energy Analysis of Biodiesel Production From Palm Oil in Thailand. Applied Energy 86, S209–S214 (2009)
12. Chin, M.J., Poh, P.E., Tey, B.T., Chan, E.S., Chin, K.L.: Biogas from Palm Oil Mill Effluent (POME): Opportunities and Challenges From Malaysia's Perspective. Renewable and Sustainable Energy Reviews 26, 717–7126 (2013)
13. Zulkifli, H., Halimah, M., Chan, K.W., Choo, Y.M., Basri, M.: Life Cycle Assessment of Oil Palm Fresh Fruit Bunch Production from Continued Land Use for Oil Palm Planted on Mineral Soil. Palm Research 22, 887–894 (2010)

14. Subramaniam, V., May, C.Y., Muhammad, H., Hashim, Z., Tan, Y.A., Wei, P.C.: Life Cycle Assessment of The Production of Crude Palm Oil. Oil Palm Research 22, 895–903 (2010)
15. Papong, S., Choom-In, T., Noksa-nga, S., Malakul, P.: Life Cycle Energy Efficiency and Potential of Biodiesel Production from Palm Oil in Thailand. Energy Policy 38, 223–226 (2010)
16. Yee, K.F., Tan, K.T., Abdullah, A.Z., Lee, K.T.: Life Cycle Assessment of Palm Biodiesel: Revealing Facts and Benefits for Sustainability. Applied Energy 86, S189–S196 (2009)
17. Harsono, S.S., Prochnow, A., Grundmann, P., Hansen, A., Hallmann, C.: Energy Balances and Greenhouse Gas Emissions of Palm Oil Biodiesel In Indonesia. GCB Bioenergy 4, 213–228 (2012)
18. Yusoff, S.: Renewable Energy from Palm Oil – Innovation on Effective Utilization of Waste. Journal of Cleaner Production 14, 87–93 (2006)
19. Yandra, Kudang, B.S., Hendra, G.: Algoritma Genetika, Teori dan Aplikasinya untuk Bisnis dan Industri. Genetic Algorithm, Application in Business and Industry. IPB Press, Bogor (2012)
20. Yandra, Marimin, Jamaran, I., Eriyatno, Tamura, H.: An Integration of Multi-Objective Genetic Algorithm and Fuzzy Logic for Optimization of Agroindustrial Supply Chain Design. In: 51st Annual Meeting of the International Society for the Systems Sciences, ISSS 2007, Tokyo, Japan (2007)

ACO-LS Algorithm for Solving No-wait Flow Shop Scheduling Problem

Ong Andre Wahyu Riyanto[1,*] and Budi Santosa[2]

[1]Department of Industrial Engineering, University of Wijaya Putra
Jl.Raya Benowo No. 1-3, Surabaya-60197, Indonesia
[2]Department of Industrial Engineering, Institut Teknologi Sepuluh Nopember,
Kampus ITS, Sukolilo, Surabaya, 60111, Indonesia
ongandre@uwp.ac.id

Abstract. In this paper, we propose a metaheuristic approach for the no-wait flow shop scheduling problem with respect to the makespan criterion. In the literature, this problem is known NP-hard type. In the literature, several algorithms have been proposed to solve this problem. We propose a hybridization of ant colony optimization (ACO) algorithm with local search (LS) in order to solve this scheduling problem, and then we call this as ACO-LS algorithm. This local search technique contributes to improve the quality of the resulting solutions. In addition, the mechanism of insert-remove technique is developed to help the searching of solution escape from the local optimum. The proposed algorithm is tested with the 31 well-known flow shop benchmark instance. The computational results based on well-known benchmarks and statistical performance comparisons are also reported. It is shown that the proposed ACO-LS algorithm is more effective than hybrid differential evolution (HDE) algorithm [Qian B., et.al, Computer & Industrial Engineering, 2009].

Keywords: Metaheuristic, No-wait flow shop scheduling, Makespan, Ant Colony Optimization, Local search.

1 Introduction

Production scheduling is one of the critical issues in the planning and manufacturing process (Pezella, et al., 2008). Scheduling problem focused on how to allocate machines to perform a collection of activities in a period of time in order to optimize a certain objective (Pinedo, 2012). In this paper, we study with the basic *n*-job *m*-machine no-wait flow shop scheduling problem. In a no-wait flow shop scheduling problem, there are *n* jobs in which each job has *m* operations and must be processed on a set of series machines continually. Once a job is started on the first machine, it must be processed through all machines without any interruption. No-wait flow shop scheduling problem is a kind of scheduling problem which has important applications including chemical processing (Rajendran, 1994), food processing (Hall and Sriskandarayah, 1996), steel production and pharmaceutical processing (Grabowski & Pempera, 2000). Given the processing time of each job in each

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 89–97, 2015.
DOI: 10.1007/978-3-662-46742-8_8

machine, the no-wait flow shop scheduling problem is to find a set of schedules to optimize a certain objective.

In this study, we consider the makespan as optimization objective. This problem is denoted as $F|prmu, no-wait|Cmax$. The no-wait flow shop scheduling problem with single objective is NP-hard (Garey and Johnson, 1979). Therefore, many researchers are interested to investigate the finding of the near optimal solution by applying heuristic and metaheuristic algorithms which can find the near optimal solution in reasonable computational time.

The no-wait flow shop scheduling problem with makespan criterion has attracted the attention by many researchers. Qian, et. al (2009) proposed HDE for the no-wait flow shop scheduling problem with the makespan criterion and showed that the solution obtained by their HDE algorithms are superior to the ones given by other previous algorithms. Laha and Sapkal (2011) proposed constructive heuristic for the no-wait flow shop scheduling problem with the total flow time criterion and show that there is a significant improvement in solution quality over the existing heuristic. Chaudhry and Mahmood (2012) proprosed a general purpose spreadsheet based genetic algorithm (GA) for solving the no-wait flow shop scheduling problem with the minimization of makespan.

In this paper we develop a hybridization between algorithm of ant colony optimization (ACO) with technique of local search (LS) in order to solve the no-wait flow shop scheduling problem with the makespan criterion. Then we call this as ACO-LS algorithm.

The rest of the paper is organised as follows: Section 2 introduces briefly the no-wait flow shop scheduling problem, followed by problem and formulation. In Section 3, ACO is briefly presented, and Section 4, gives the proposed ACO-LS details as implemented in this paper. Experimental results are given in Section 5. The last section of the paper presents conclusions.

2 No Wait Flow Shop Scheduling Problem

The no-wait flow shop scheduling problem can be described as follows: given the processing time $p(i,j)$ of job i on machine j, each of n jobs will be sequentially processed on machine 1, 2, ..., m. At any time, each machine can process at most one job and each job can be processed on at most one machine. The sequence in which the jobs are to be processed is the same for each machine. To meet the no-wait restrictions, the completion time of a job on given machine must be equal to the starting time of the job on the next machine. In other words, there must be no waiting time between the processing of any consecutive operations of each n jobs. The problem is to find a sequence that the given criteria, i.e., makespan is optimized.

Let $= (\pi_1, \pi_2, \ \pi_n)$ denote the schedule or permutation of jobs to be processed, $L(\pi_{j-1}, \pi_j)$ the minimum delay on the first machine between the start of job π_j and π_{j-1} restricted by the no-wait constraint. Then L can be calculated as follows (Qian, et. al, 2009):

$$L(\pi_{j-1}, \pi_j) = p(\pi_{j-1}, 1) + max\left[0, \underset{2\le k\le m}{max}\left\{\sum_{h=2}^{k} p(\pi_{j-1}, h) - \sum_{h=1}^{k-1} p(\pi_j, h)\right\}\right] \quad (1)$$

Thus, the makespan can defined as follows :

$$C_{max} = \sum_{j=1}^{n} L(\pi_{j-1}, \pi_j) + P_{sum}(\pi_n) \quad (2)$$

where $P_{sum}(\pi_n) = \sum_{k=1}^{m} p(\pi_n, k)$

Fig. 1 shows an example of no-wait flow shop scheduling problem with n=4 and m=4

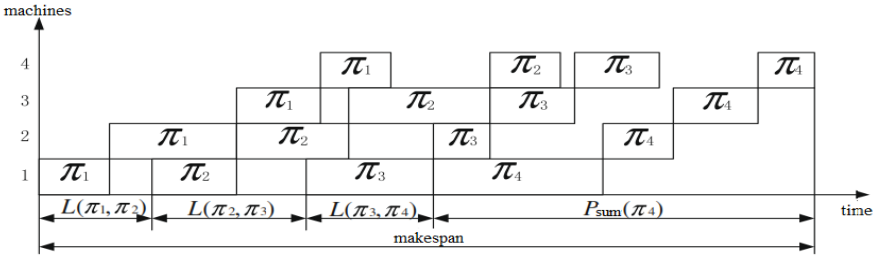


Fig. 1. No-wait Flow Shop Scheduling Problem with n=4 and m=4

3 Structure of Ant Colony Optimization

Ant Colony Optimization (ACO) is a class of metaheuristic approach proposed by Marco Dorigo in the early 1990s. The main idea ACO is to mimic the pheromone trail used by real ants for finding a nearest trail between their nest and the food sources, as shown in figure 2. Ant communication is accomplished through chemicals that called pheromones. Ants communicate to one another by laying down pheromones along their trails. Other ants perceive the presence of pheromone and tend to follow paths where pheromone concentration is higher. New pheromone will be released on the chosen path, which makes it more attractive for next ants. Shortly, all ants will select the shortest path. This analogy was applied to find the best solution to the problem of combinatorial optimization such as scheduling problem.

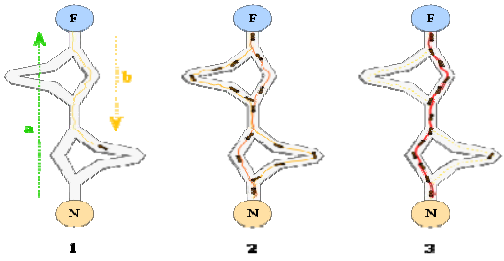


Fig. 2. Basic ant behavior: 1)Ants in a pheromone trail between nest and food. 2) Ants find paths to go around the obstacle. 3) A new pheromone trail is formed along the shorter path.

The general procedure of ACO algorithm is described as follows :

3.1 Heuristic Information

The heuristic information value is also initialized at the initialization step. The use of heuristic information to direct the ants' probabilistic solution construction is important because it provide problem specific knowledge. Heuristic information used in this study is the distance between two jobs SPIRIT (Sequencing Problem Involving a Resolution by Integrated Taboo Search Techniques) rule presented by Widmer and Hertz (1989). We modify SPIRIT method for no-wait flow shop scheduling problem. The distance between the start of job π_j and π_{j-1} , $d(\pi_{j-1}, \pi_j)$ is given by the following equation:

$$d(\pi_{j-1}, \pi_j) = L(\pi_{j-1}, \pi_j) \tag{3}$$

$$\eta(\pi_{j-1}, \pi_j) = 1/d(\pi_{j-1}, \pi_j) \tag{4}$$

3.2 Solution Construction

In the iterative step, an ants colony determines starting jobs. Each ant repeatedly applies the state transition rule to select the next processing job up to a complete schedule is formed. When building a schedule, both the heuristic information and pheromone amount are used to choose the next job. While constructing the schedule, an ant also decreases the amount of pheromone between selected jobs by applying the local updating rule to vary other ants schedule and to avoid in leading to local optima.

While finding appropriate solution, after the ant *k* chooses the next job to move to by applying the state transition rule, selected job is added into tabu list. Until the last job is selected, the procedure is repeated.

3.3 State Transition Rule

In the process of schedule constructed, the ant *k* in job *i* selects the job *j* ($i \neq j$) to move by applying the following state transition rule:

$$j = \begin{cases} \arg\max_{u \in S_k(i)} \left\{ [\tau(i,u)]^{\alpha} \left[\eta(i,u) \right]^{\beta} \right\} & \text{if } q \leq q_0 \\ J & \text{otherwise} \end{cases} \tag{5}$$

where, $\tau(i,u)$ is the amount of pheromone trail on edge (i,u). $Where\ \eta(i,u) = 1/\delta(i,u)$ is the inverse of the distance $\delta(i,u)$ between job *i* and job *u* denotes the reciprocal of a cost measure between nodes *i* and *u*. In the no-wait flow shop scheduling problem, $\delta(i,u)$ is identical with $d(\pi_{j-1}, \pi_j)$. $S_k(i)$ is the set of feasible jobs to be selected by ant *k* in job *i*. It is clear that the set of feasible jobs not contained in tabu. α is a parameter that allow a user to control the relative importance of pheromone trail $(\alpha > 0)$. β is a parameter that determines the relative importance of heuristic information. $(\beta > 0)$, q is a value chosen randomly with uniform probability in

[0,1] and q_0 is a parameter that determines the relative importance of exploitation versus exploration $(0 \leq q_0 \leq 1)$. J is a random variable selected according to the following random-proportional rule probability distribution, which is the probability with that ant k chooses to move from job i to job j:

$$p_{k(i,j)} = \begin{cases} \frac{[\tau(i,j)]^{\alpha}\left[\eta(i,j)\right]^{\beta}}{\sum_{u \in S_k(i)}[\tau(i,u)]^{\alpha}\left[\eta(i,u)\right]^{\beta}} & \text{if } j \in S_k(i) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

3.4 Pheromone Trail Update

While an ant construct a schedule, an ant decreases the level of pheromone trail between selected jobs by applying the update rule. The update rules consists of two terms: the first, is the evaporation of the existing pheromone (local updating rule) ; the second, is the amount of added pheromone on the trail (global updating rule). The local updating rule is formulated as follows :

$$\tau(i,j) = (1 - \rho l) \cdot \tau(i,j) + \rho l \cdot \tau_0 \quad (7)$$

where, τ_0 is the initial pheromone level and ρl $(0 < \rho l < 1)$ is the local pheromone evaporation parameter.

After all ants completed their schedules then global updating rule is performed. Global updating rule provides a greater amount of pheromone trail for between adjacent jobs of the best schedule. The pheromone trail level is updated as follows:

$$\tau(i,j) = (1 - \rho g) \cdot \tau(i,j) + \rho g \cdot \Delta\tau(i,j) \quad (8)$$

where,

$$\Delta\tau(i,j) = \begin{cases} (L_b)^{-1} & if\ (i,j) \epsilon \text{ best schedule} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

In the Eq. (8), $\rho g (0 < \rho g < 1)$ is the evaporation parameter of global updating rule and L_b is the objective function value of the best schedule until the current iteration.

4 Proposed Ant Colony Optimization – Local Search (ACO-LS)

The main advantages such as ACO metaheuristic approach is able to find the near optimal solution in reasonable computational time.

The structure of proposed ACO can be describe as follows :

1. **Initialize** the pheromone trails, heuristic information, and parameters
2. **Iteration :**
   2.1 Ant colonies determine the starting jobs ;
   2.2 Each ant constructs a complete solution ;

Repeat
Applying state transition rule for selecting the next jobs
Until complete jobs

2.3 Improve the solution by local search

3. **Cycle**. If the maximum number of iterations is reached, then the iteration stops. If not then go back to step 2.
4. Return best solution found

Local Search

ACO algorithm can perform better to find solution when combined with a local search algorithm (Yagmahan and Yenisey, 2010). In this study, we propose a local search procedure as follows:

Step 1:

Determine one best solution of job sequence π_{i_0}

Step 2:

Choose randomly u and v, where $u \neq v$; $\pi_i = insert(\pi_i, u, v)$.

Step 3:

Set loop=0

While $loop < n * (n-1)$

$count = 0$;

max_$method = 3$;

While $count <$ max _$method$

Choose randomly u and v, $u \neq v$;

$\pi_{i\_1} = insert(\pi_i, u, v)$;

if $f(\pi_{i\_1}) < f(\pi_i)$,

then $\pi_i = \pi_{i\_1}$; $count = 0$;

if $f(\pi_{i\_1}) > f(\pi_i)$,

then $\pi_i = \pi_i$; $count = count + 1$;

$choose\ randomly \quad u, v, r, s$; $u \neq v \neq r \neq s$;

$\pi_{i\_1} = remove(\pi_i, u, v) \rightarrow remove(\pi_i, r, s)$;

if$(\pi_{i_1}) < f(\pi_i)$,

then $\pi_i = \pi_{i\_1}$; $count = 0$;

if $f(\pi_{i\_1}) > f(\pi_i)$,

then $\pi_i = \pi_i$; $count = count + 1$;

choose randomly t, u, v, q, r, s; t≠ u≠ $v \neq q \neq$ r≠ s;

$\pi_{i\_1} = remove(\pi_i, t, u, v) \rightarrow remove(\pi_i, q, r, s)$;

if $f(\pi_{i\_1}) < f(\pi_i)$,

then $\pi_i = \pi_{i\_1}$; $count = 0$;

if $f(\pi_{i\_1}) > f(\pi_i)$,

then $\pi_i = \pi_i$; $count = count + 1$;

end

end

Step 3:

if $f(\pi_i) < f(\pi_{i\_0})$,

then $\pi_{i\_0} = \pi_i$;

if $f(\pi_i) < f(\pi_{i\_0})$,

then $\pi_{i\_0} = \pi_{i\_0}$;

Where : n is set of all the job. π_i is sequence of n job. While t, u, v, q, r, s is the position of a job in the sequence π_i .

Termination Criteria

Termination criteria using the maximum number of iterations

Table 1. Comparison of HDE and ACO-LS

		RAJ	HDE			ACO-LS		
Problem	**n,m**	**C***	**BRE**	**ARE**	**WRE**	**BRE**	**ARE**	**WRE**
Car1	11,5	8142	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.27
Car2	13,4	8242	**0.00**	**0.00**	0.06	**0.00**	**0.00**	**0.00**
Car3	12,5	8866	**0.00**	0.03	**0.08**	**0.00**	**0.00**	**0.08**
Car4	14,4	9195	**0.00**	1.15	2.41	**0.00**	**0.06**	**0.68**
Car5	10,6	9159	**0.00**	**0.00**	**0.00**	0.39	0.64	2.23
Car6	8,9	9690	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Car7	7,7	7705	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Car8	8,8	9372	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Rec01	20,5	1590	-3.71	-3.50	-3.14	**-4.03**	**-3.88**	**-3.83**
Rec03	20,5	1457	**-6.59**	-5.33	-3.77	**-6.59**	**-6.23**	**-4.99**
Rec05	20,5	1637	**-7.70**	-6.76	-5.86	-7.64	**-7.35**	**-7.20**
Rec07	20,10	2119	-3.59	-3.19	-1.27	**-3.63**	**-3.55**	**-3.44**
Rec09	20,10	2141	-4.58	-4.30	-3.60	**-4.62**	**-4.56**	**-4.31**
Rec11	20,10	1946	**-3.34**	**-3.00**	**-2.21**	**-3.34**	-2.90	**-2.21**
Rec13	20,15	2709	-5.76	-4.72	-3.32	**-6.05**	**-5.73**	**-4.89**
Rec15	20,15	2691	**-6.02**	-5.89	-5.39	**-6.02**	**-5.98**	**-6.25**
Rec17	20,15	2740	**-5.58**	**-5.55**	**-5.47**	**-5.58**	-5.50	-5.44
Rec19	30,10	3157	-7.73	-6.76	-6.34	**-9.72**	**-8.78**	**-8.01**
Rec21	30,10	3015	-4.61	-3.74	-3.32	**-6.27**	**-6.10**	**-6.07**
Rec23	30,10	3030	-8.71	-7.69	-7.19	**-10.89**	**-10.16**	**-9.92**
Rec25	30,15	3835	-4.64	-4.02	-3.60	**-6.31**	**-5.67**	**-5.44**
Rec27	30,15	3655	-3.97	-3.07	-2.71	**-6.10**	**-5.39**	**-4.89**
Rec29	30,15	3583	-6.61	-5.98	-5.44	**-8.15**	**-7.60**	**-7.23**
Rec31	50,10	4631	-3.95	-3.22	-2.92	**-5.92**	**-4.95**	**-3.67**
Rec33	50,10	4770	-2.37	-1.82	**-1.26**	**-3.86**	**-2.48**	-0.76
Rec35	50,10	4718	-3.98	-3.77	**-3.39**	**-5.26**	**-3.86**	-1.82
Rec37	75,20	8979	-7.57	-7.25	**-6.97**	**-8.56**	**-7.39**	-6.57
Rec39	75,20	9158	-3.53	-3.07	-2.86	**-5.43**	**-4.36**	**-2.89**
Rec41	75,20	9344	-5.15	-4.84	-4.54	**-6.93**	**-6.09**	**-4.89**
Hel1	100,10	780	-5.26	-4.73	-4.36	**-6.92**	**-5.81**	**-5.09**
Hel2	20,10	189	**-5.29**	-4.23	-1.59	**-5.29**	**-4.92**	**-4.30**
Average			-3.88	-3.40	-2.84	**-4.60**	**-4.15**	**-3.58**

5 Experimental Result

In this section, the results of computational experiments performed are presented to evaluate the performance of proposed ACO-LS. The performance of the proposed ACO-LS is tested with numerical simulations are carried out with 31 well-studied benchmark contributed to the OR-Library (http://www.people.brunel.ac.uk/-mastjjb/jeb/info.html). The first 8 problems are called Car1, Car 2 through Car8 by Carlier (1978). The second 21 problem are called Rec01, Rec03 through Rec41by Reeves (1995). The last two problem Hel1 and Hel2 by Heller (1960). The performance of ACO-LS using test problem is compared with HDE algorithm proposed by Qian, et. al (2009). All algorithms are coded in matlab 2009a and run on PC with 2 GHz Intel Core 2 Duo processor and 2 GB RAM memory. For fair comparison, HDE algorithm are executed on the same PC and let HDE algorithm run at the same time as ACO-LS. Each benchmark is independently run 20 time for comparison. Where C* denotes the references makespan produced by the famous RAJ heuristic (Rajendran, 1994). BRE denotes the best relative percentage error to C*. ARE denotes the average relative percentage error to C*, and WRE denotes the worst relative percentage error to C*. The statistical results are reported in table 1.

It can be seen from table 1 that the BRE, ARE, and WRE values performed by ACO-LS are much better than those obtained by HDE almost for all benchmark.

6 Conclusion

In this paper, we have presented an ACO-LS algorithm for the no-wait flow shop scheduling problem with the objective of minimizing makespan criterion. Based on the computational experimentation, the proposed ACO-LS algorithm gives comparable performance as that of HDE algorithm for small problem sizes, whereas, there is significant improvement in solution quality for large problem sizes. To the best our knowledge, this is the first report to apply ACO-LS algorithm for no-wait flow shop scheduling problems with makespan. In future work, we will extend the ACO-LS algorithm to the scheduling problem with two or more objectives.

Acknowledgments. This research is funded by DP2M Ditjen Dikti with the grants scheme of Pekerti (Nomor: 009/SP2H/P/K7/KM/2014, 19 May 2014). We would like to thank you for DP2M Ditjen Dikti, Kopertis 7, and LPPM-UWP.

References

1. Chaudhry, I.A., Mahmood, S.: No-wait Flowshop Scheduling Using Genetic Algorithm. In: Proceedings of the World Congress on Engineering, vol. 3 (2012)
2. Carlier, J.: Ordonnancements a contraintes disjonctives. RAIRO-Operations Research-Recherche Opérationnelle 12(4), 333–350 (1978)
3. Grabowski, J., Pempera, J.: Sequencing of jobs in some production system. European Journal of Operational Research 125(3), 535–550 (2000)

4. Garey, M.R., Johnson, D.S.: Computers and intractability, aguide to the theory of NP-completeness. Freeman, San Francisco (1979)
5. Hall, N.G., Sriskandarajah, C.: A survey of machine scheduling problems with blocking and no-wait in process. Operations Research 44(3), 510–525 (1996)
6. Heller, J.: Some numerical experiments for an M× J flow shop and its decision-theoretical aspects. Operations Research 8(2), 178–184 (1960)
7. Laha, D., Sapkal, S.U.: An efficient heuristic algorithm for m-machine no-wait flow shops. In: Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1 (2011)
8. Pinedo, M.L.: Scheduling: theory, algorithms, and systems. Springer (2012)
9. Pezzella, F., Morganti, G., Ciaschetti, G.: A genetic algorithm for the flexible job-shop scheduling problem. Computers & Operations Research 35(10), 3202–3212 (2008)
10. Qian, B., Wang, L., Hu, R., Huang, D.X., Wang, X.: A DE-based approach to no-wait flow-shop scheduling. Computers & Industrial Engineering 57(3), 787–805 (2009)
11. Rajendran, C.: A no-wait flowshop scheduling heuristic to minimize makespan. Journal of the Operational Research Society, 472–478 (1994)
12. Reeves, C.R.: A genetic algorithm for flowshop sequencing. Computers & Operations Research 22(1), 5–13 (1995)
13. Rajendran, C.: A no-wait flowshop scheduling heuristic to minimize makespan. Journal of the Operational Research Society, 472–478 (1994)
14. Widmer, M., Hertz, A.: A new heuristic method for the flow shop sequencing problem. European Journal of Operational Research 41(2), 186–193 (1989)
15. Yagmahan, B., Yenisey, M.M.: A multi-objective ant colony system algorithm for flow shop scheduling problem. Expert Systems with Applications 37(2), 1361–1368 (2010)

A New Ant-Based Approach for Optimal Service Selection with E2E QoS Constraints

Dac-Nhuong Le[1] and Gia Nhu Nguyen[2]

[1] Faculty of Information Technology, Haiphong University, Haiphong, Vietnam
[2] Duytan University, Danang, Vietnam
Nhuongld@hus.edu.vn, Nguyengianhu@duytan.edu.vn

Abstract. In this paper, we study the dynamic service composition becomes a decision problem on which component services should be selected under the E2E QoS constraints. This problem is modeled is a nonlinear optimization problem under several constraints. We have two approaches: local selection is an efficient strategy with polynomial time complexity but can not handle global constraints and the traditional global selection approach based on integer linear programming can handle global constraints, but its poor performance renders it inappropriate for practical applications with dynamic requirements. To overcome this disadvantage,, we proposed a novel Min-Max Ant System algorithm with the utilities heuristic information to search the global approximate optimal. The experiment results show that our algorithm is efficiency and feasibility more than the recently proposed related algorithms.

Keywords: E2E QoS constraints, QoS-Aware, Min-Max Ant System.

1 Introduction

Nowadays, the web services technologies are new solution to develop software foundation for next generation enterprise and Web-based systems. Services from different providers can be integrated into a composite service regardless of their locations, platforms, and/or execution speeds to implement complex business processes and transactions that can be provide a promising solution for the integration of business application [1]. With growing demands, many service providers have started to offer different QoS (*Quality of service*) service levels so as to meet the needs of different user groups, by offering different service qualities based on user needs or service cost [2]. For composite services, one of the QoS issues is to define the service integration model and identify the optimal service selection to meet a users QoS requirement. Component services can be dynamically composed to form complex and value-added services so that the requirement of users be satisfied. Therefore, the end-to-end (E2E) services selection is an important part of the web service composition problem [3].

In this paper, we study the dynamic service composition becomes a decision problem on which component services should be selected under the E2E QoS constraints. This problem is modeled is a decision problem as a nonlinear optimization problem under several constraints. The rest of this paper is organized as

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 98–109, 2015.
DOI: 10.1007/978-3-662-46742-8_9

follows: some related works will be analyzed and evaluated in section 2. Section 3 introduces the model system defined for service selection with E2E QoS constraints problem. Section 4 presents our new an ant-based approach to solve it. Section 5 is our simulation and analysis results based on performance compared, and finally, section 6 concludes the paper.

2 Related Works

There are many have worked on this topic, such as: BPEL4WS (*Business Process Execution Language for Web services*), Microsofts XLANG, IBMs WSFL (*Web service Flow Language*) , WSCI (*Web service Choreography Interface*), BPML (*Business Process Modeling Language*), DAML-S (*DARPA Agent Markup Language*) are the industrial standard specifications for Web service composition. The SWORD project, FUSION framework, eFlow system, QoS-oriented framework WebQ, SELF-SERV platform are the centralized operator, engine, coordinator to control the execution of constructed composite services [1]. An quality driven approach to select component services during execution of a composite service presented in [4], the authors consider multiple QoS criteria such as price, duration, reliability and take into account of global constraints. The linear programming method is used to solve the service selection problem, which is usually too complex for run-time decisions. The E2E QoS model and management architecture for complex distributed real-time system [5].

The QoS-based service selection problem aims at finding the best combination of web services that satisfy a set of E2E QoS constraints in order to fulfill a given SLA (*Service Level Agreement*), which is an NP-Hard problem [6]. This problem becomes especially important and challenging as the number of functionally equivalent services offered on the web at different QoS levels increases exponentially [7]. As the number of possible combinations can be very huge, based on the number of subtasks comprising the composite process and the number of alternative services for each subtask. The local selection strategy in which, the candidate service who has the maximum QoS utility value is selected for each service class is very efficient but can not handle end-to-end QoS constraints [9]. D.Ardagna et al proposed exact search algorithms to perform an exhaustive search to find the best combination that satisfies a certain composition level SLA is impractical [8]. T.Yu et al proposed heuristic algorithms that can be used to find a near-optimal solution efficiently in [3]. M.Wu et al proposed the global optimization algorithm uses integer linear programming techniques to find the optimal selection of component services for the composition [10]. However, the disadvantage of deterministic algorithms like linear programming are only applicable to small-scale problems due to their exponential time complexity. In [11], C.W.Zhang et al are proposed a *Genetic Algorithm* (DiGA) to solve the web service selection supporting QoS. Similar, S.Liu et al in [12] are applied genetic algorithm for dynamic web services selection algorithm with QoS global optimal in web services composition. The advantages of genetic algorithm are intelligence and global convergence. Genetic algorithm which is inspired from the evolution of

biological populations is a non-deterministic algorithm. Although very efficient in solving NP-Hard problem like service selection, the effect of genetic algorithm is very sensitive to its parameter setting. The *Parameter-Adaptive Genetic Algorithm* (PAGA) solved the service selection problem presented in [13]. In [14], K.Su et al suggested an efficient discrete invasive weed optimization algorithm for web services selection consists two steps. First, a set of randomly generated feasible solutions are transformed into decimal code. Second, they utilize Gaussian diffusion to guide the population tospread in the solution space. X.Q.Fan et al is presented a new cooperative evolution algorithm consisting of a *Stochastic Particle Swarm Optimization* (SPSO) [15] and a *Simulated Annealing* (SA). The *Multipheromone and Dynamically Updating Ant Colony Optimization* algorithm (MPDACO) put forward to solve this problem, which includes a global optimization process and a local optimizing process proposed by Y.M.Xia et al [16]. C.Zhange et al proposed the *Clustering Based Ant Colony Algorithm for Service Selection* (CASS) including its model and concrete algorithm in [17].

3 Problem Formulation

3.1 Notations and Definitions

The concepts definition for service selection with E2E QoS constraints problem [18] can be defined as as follows:

1. *Service Class*: service class $S_j = \{s_{1j}, s_{2j}, ..., s_{mj}\}$ is used to describe a set of functionality equivalent web services, where $s_{ij}(1 \leq i \leq m)$ represents the i^th component service in service class S_j.
2. *Abstract Composite Service*: an abstract composite service $CS_{abstract} = \{S_1, S_2, ..., S_n\}$ can be defined as an abstract representation of a composition request. Each task of the abstract composite service refers to a service class without referring to any concrete service.
3. *Concrete composite service*: A concrete composite service CS is a instantiation of an abstract composite service. It can be gained by binding each service class $S_j \in CS_{abstract}$ to a concrete service s_{ij}.
4. *QoS criteria*: which can be used to describe the quality of web services, is a set of non-functional properties. QoS criteria can be divided into several aspects such as runtime-related QoS, transaction support related QoS, configuration management QoS, cost related QoS and security related QoS [19]. More generally, QoS criteria can include generic QoS like response time, price, reliability, availability etc, as well as domain specific QoS [18]

An extensible QoS computation model includes both the generic and domain specific criteria, where new criteria can be added and used to evaluate the QoS without changing the underlying model are presented in [18]. The QoS criteria of a concrete service s can be defined as a vector $Q(s) = [q_1(s), q_2(s), ..., q_r(s)]$ where $q_k(s)(1 \leq k \leq r)$ represents the k^{th} QoS criterion of s. The QoS criteria of a concrete composite service CS can be defined as a vector $Q(CS) = [q'_1(CS), q'_2(CS), ..., q'_k(CS)]$, where $q'_k(CS)(1 \leq k \leq r)$ represents the k^th QoS criterion of CS [18].

3.2 Web Service QoS Attributes

The QoS attributes of a business process are decided by the QoS attributes of individual services and their composition relationships. There are different ways that individual services can be integrated to form a business process. The QoS of a concrete composite service is decided by the QoS of its component services and the composition model. The four basic relationships are sequential, parallel, conditional and loop. Our composition model that we focus in this paper is sequential, due to the other models can be transformed to the sequential model by using some technologies like *Critical Path Algorithm*, *Hot Path*, and *Loop Unfolding* [4].

In this study, we are considered four quality attributes as part of the Web service parameters: *response time, reliability, availability and throughput*. The response time of the business process is the sum each individual service s_j response time at the chosen service level. The reliability of the business process is the product of each individual service s_j reliability at the selected service level. It is a non-linear function. In order to make all aggregation functions to be linear, we can transform it using a logarithmic function. The availability of the business process is the product of each individual service s_j availability at the selected service level. Same as the reliability attribute, we can use an logarithmic function to convert it to a linear formulation. The computation expression of the QoS of composite services are shown in Table 1.

Table 1. Computation expression of the QoS of composite services

QoS Criteria	Aggregation Function
Response Time	$q'_{time}(CS) = \sum_{j=1}^{n} q_{time}(s_j)$
Reliability	$q'_{rel}(CS) = \prod_{j=1}^{n} q_{rel}(s_j)$
Availability	$q'_{av}(CS) = \prod_{j=1}^{n} q_{av}(s_j)$
Throughput	$q'_{thr}(CS) = \min_{1 \le j \le n} q_{thr}(s_j)$

Criteria like reliability, availability and throughput are positive that means the higher the value is, the higher the quality is. On the contrary, response time is a negative criterion that means the higher the value is, the lower the quality is. We utilize the *Simple Additive Weighting* (SAW) technique to map the QoS vector into a single real value for the purpose of evaluating the quality of composite services [4]. The utility of composite service CS is computed as:

$$U(CS) = \sum_{k=1}^{1} w_k \frac{Q_{\max}(k) - q'_k(CS)}{Q_{\max}(k) - Q_{\min}(k)} + \sum_{k=2}^{4} w_k \frac{q'_k(CS) - Q_{\min}(k)}{Q_{\max}(k) - Q_{\min}(k)} \quad (1)$$

In which, $Q_{min}(k)$ and $Q_{max}(k)$ represent the minimum and maximum value of the k^{th} QoS criterion of all possible instantiations of the given abstract

composite service can be computed as follows:

$$Q_{\min}(k) = \sum_{j=1}^{n} \min_{s_{ij} \in S} q_k(s_{ij}) \text{ and } Q_{\max}(k) = \sum_{j=1}^{n} \max_{s_{ij} \in S} q_k(s_{ij}) \tag{2}$$

and $w_k (1 \leq k \leq 4)$ is the weight assigned to each QoS criterion, which represents users priorities and following constraints:

$$\sum_{k=1}^{4} w_k = 1, 0 \leq w_k \leq 1, w_k \in R \tag{3}$$

3.3 QoS-Aware Service Composition

The process of QoS-Aware service composition can be defined as follows:

1. An abstract composite service can be stated in a workflow-like language like *Process Execution Language for Web Services* (BPEL4WS).
2. The discovery engine utilizes existing infrastructure like *Universal Description Discovery and Integration* (UDDI) or Seekda [20] to discover several concrete services for each task in the work-flow by means of syntactic or semantic matching.
3. The service selection middle-ware selects one component service for each task. The component services selected should maximize the utility of composite service without violating the E2E QoS constraints

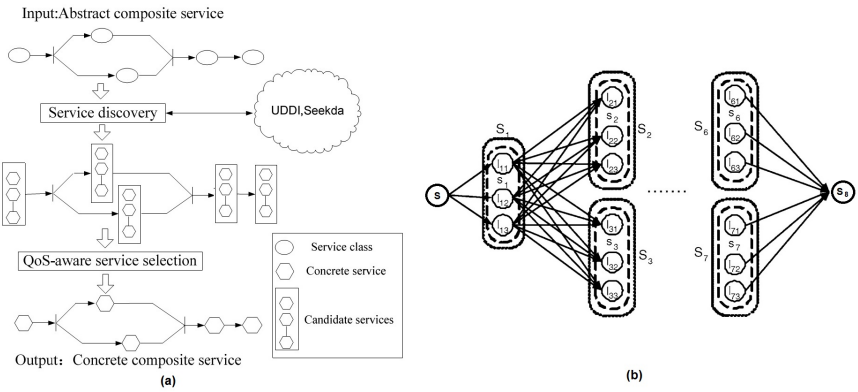


Fig. 1. QoS-Aware service composition and Graph construction

We use binary decision variables x_{ij} to represent whether the candidate service s_{ij} is selected or not ($x_{ij} = 1$ means that s_{ij} is selected, and otherwise). We can rewriting formula (1) to include the decision variables and the QoS computation expression of composite services defined in Table 1.

The problem of QoS-Aware service selection (QASS) can be formulated as a maximization problem of the utility value [16] given by:

$$F = \max \left(\begin{array}{l} w_1 \frac{Q_{\max}(1) - \sum_{j=1}^{n} \sum_{i=1}^{m} q_1(s_{ij}) x_{ij}}{Q_{\max}(1) - Q_{\min}(1)} + \sum_{k=2}^{3} w_k \frac{\prod_{j=1}^{n} \sum_{i=1}^{m} q_k(s_{ij}) x_{ij} - Q_{\min}(k)}{Q_{\max}(k) - Q_{\min}(k)} \\ \qquad + w_4 \frac{\min_j \left(\sum_{i=1}^{m} q_4(s_{ij}) x_{ij} \right) - Q_{\min}(4)}{Q_{\max}(4) - Q_{\min}(4)} \end{array} \right) \tag{4}$$

Subject to:

$$\sum_{i=1}^{m} x_{ij} = 1, 1 \leq k \leq n \tag{5}$$

$$\sum_{j=1}^{n} \sum_{i=1}^{m} q_1 (s_{ij}) x_{ij} \leq C_1 \tag{6}$$

$$\prod_{j=1}^{n} \sum_{i=1}^{m} q_k (s_{ij}) x_{ij} \leq C_k, k = 2, 3 \tag{7}$$

$$\min_j \left(\sum_{i=1}^{m} q_4 (s_{ij}) x_{ij} \right) \leq C_4 \tag{8}$$

where C_1, C_2, C_3, C_4 represent the E2E constraints on response time, reliability, availability and throughput respectively. Constraint (5) means that only one concrete service should be selected for each task. From the above discussion, we can find that The QASS problem is a nonlinear optimization problem under several constraints. Our approach for this problem is presented in the next section. This problem is proved to be a NP-Complete problem [13]. Deterministic algorithms like linear programming are only applicable to small-scale problems due to their exponential time complexity.

4 Our Approach

The *Min-Max Ant System* (MMAS) based on several modifications to *Ant System* (AS) which aim (*i*) to exploit more strongly the best solutions found during the search and to direct the ants search towards very high quality solutions and (*ii*) to avoid premature convergence of the ants' search [21].

4.1 Construct Ant Solutions

In the first step, we generated the construct graph for services. Each service level in every individual service is represented by a node in the graph, with a cost and a benefit value in Fig.1.

1. Each service level of each individual service is represented as a node in the graph.

2. If service s_i is connected to service s_j, all service levels in s_i are connected to all service levels in s_j.
3. Set link cost, delay and benefit: To remove parameters from graph nodes, we add the service time and cost of the receiving node to all incoming links to the node. That is, if data is sent from s_i to s_j, we will add service time $e(s_j, l)$ and cost $c(s_j, l)$ to the link connected to s_j. So for the link from service level l_a of s_i to service level l_b of s_j , its delay is set to $d_{ij} + e(s_j, l_b)$ and its cost is $c_{ij} + c(s_j, l_b)$. The benefit of the link is set to $b(s_j, l_b)$.
4. Suppose there are k service classes in the execution plan, we add a source node (S_0) and a sink node ($S_k + 1$) to the graph as shown in Fig.1. For all nodes that have no incoming edges, add links from the source node to them; the delay, cost and benefit of these links are set to 0. For all vertexes that have no outgoing edges, connect them with the sink node. The delay and cost of these links are set to be the transmission delay and cost between the service (*that provides the start node of the link*) and user. The benefit of these links are set to 0.
5. For each link, according to formula (4), we can compute its utility F based on the benefit and cost of the link.

4.2 Update Pheromones

Let $\tau_{\min}$ and $\tau_{\max}$ are lower and upper bounds on pheromone values. We initialize all pheromone trails to $\tau_{\max}$ and choose first service s_i has been randomly chosen. For each candidate service s_j, the pheromone factor $\tau_{S_k}(s_j)$ is initialized to $\tau(s_i, s_j)$. Each time a new service s_l is added to the solution S_k, for each candidate service s_j , the pheromone factor $\tau_{S_k}(s_j)$ incremented by $\tau(s_l, s_j)$. And the highest pheromone factor can be defined as

$$\tau_{S_k}(s_{ij}) = \sum_{s_j \in S_k} \tau(s_i, s_j) \tag{9}$$

When the algorithm starts, ants can perform search in order of group that impose the lowest restriction on the next choices. The key point is to decide which components of the constructed solutions should be rewarded, and how to exploit these rewards when constructing new solutions. A solution is a set of selected objects $S = \{s_{ij} | x_{ij} = 1\}$. Say $x_{ij} = 1$ is the mean service s_{ij} is selected and the corresponding decision variable x_{ij} has been set to 1. We must constructed solutions $S = \{x_{i_1 j_1}, ..., x_{i_n j_n}\}$, pheromone trails are laid on each objects selected in S. So pheromone trail τ_{ij} will be associated with object s_{ij}.

Let S_k be the set of the selected objects at the k^{th} iteration. The heuristic factor $S_k(s_{ij}) = q_k(s_{ij})\, x_{ij}$. We have unique idea of a separate pheromone trail for services to save the ordering of services that lead to a good solution. The service pheromone trail also follow a MMAS approach and initialized to the max pheromone value. Once each ant has constructed a solution, pheromone trails laying on the solution objects are updated according to the ACO [22]. First, all amounts are decreased in order to simulate evaporation. This is done by

multiplying the quantity of pheromone laying on each object by a pheromone persistence rate $(1 - P)$ such that $0 \leq P \leq 1$. Then, pheromone is increased for all the objects in the best solution of the iteration. When constructing a solution, an ant starts with an empty solution. At the first construction step an ant selects a group randomly and at all the latter steps, groups are selected according to their associated pheromone value. After selecting a group, the algorithm removes all the bad *Candidates* that violates resource constraints. The local heuristic information of the remaining candidate objects of the services and selects an object updated following probability equation:

$$P_{S_k(s_{ij})} = \frac{[\tau_{S_k}(s_{ij})]^{\alpha}[\eta_{S_k}(s_{ij})]^{\beta}}{\sum_{s_{ij} \in Candidates} [\tau_{S_k}(s_{ij})]^{\alpha}[\eta_{S_k}(s_{ij})]^{\beta}} \tag{10}$$

in which, *Candidates* are all items from the currently selected service which do not violate any resource constraints; τ_{S_k} is the pheromone factor of the dynamic heuristic information S_k; η_{S_k} is the desirability of s_{ij}.

$$\eta_{S_k}(s_{ij}) = \frac{1}{q_k(s_{ij})\, x_{ij}} \tag{11}$$

The influence of the pheromone concentration to the probability value is presented by the constant α, while constant β do the same for the desirability and control the relative importance of pheromone trail versus local heuristic value.

$$\tau_{S_k}(s_{ij}) \leftarrow (1 - \rho)\tau_{S_k}(s_{ij}) + \Delta\tau_{S_k}(s_{ij})^{best} \tag{12}$$

Let $S_{internalbest}$ be the best solution constructed during the current cycle. The quantity of pheromone increased for each object is defined by

$$G(S_{internalbest}) = \sum_{s_{ij} \in S_{internalbest}} \frac{1}{\sum_{j=1}^{n}\sum_{i=1}^{m} q(s_{ij})x_{ij}} \tag{13}$$

After each iteration, service pheromone values have much better than the original pheromone trail for ants. The best solution updates the group pheromone trail. All the adjacent groups get the highest amount of pheromone value that gradually diminishes as the distance between services increases. The *Candidate-groups* data structure maintains a list of feasible candidate groups which can be considered next. After each ant has constructed a solution, the best solution of that iteration is identified and a random local search procedure and a random item swap procedure is applied to improve it. Then pheromone trail is updated according to the best solution. Also it maintains a database of top k solutions. After each iteration a small amount of pheromone is deposited in the pheromone trails of the objects belonging to the top k solutions. The motivation behind this strategy is to ensure quick convergence on good solutions and to explore better areas more thoroughly. The algorithm stops either when an ant has found an optimal solution, or when a maximum number of cycles has been performed. After pheromone update, $\tau_{S_k}(s_{ij})$ is set in $[\tau_{min}, \tau_{max}], \forall s_{ij}$ defined by:

$$\tau_{S_k}(s_{ij}) = \begin{cases} \tau_{\max} \text{ if } \tau_{S_k}(s_{ij}) > \tau_{max} \\ \tau_{\max} \text{ if } \tau_{S_k}(s_{ij}) \in [\tau_{min}, \tau_{max}] \\ \tau_{\min} \text{ if } \tau_{S_k}(s_{ij}) < \tau_{min} \end{cases} \quad (14)$$

4.3 Our Algorithm

Algorithm 1. MMAS algorithm for Service Selection with E2E QoS Constraints

Initialize pheromone trails to τ_{max} for both item and group;
$Top_{ksolution} \Leftarrow$ {hold topmost k solutions};
Repeat
  $S_{globalbest} \Leftarrow \oslash$;
  For each ant $k = 1...N_{Ant}$ **do**
    $S_{internalbest} \Leftarrow \oslash$;
    Candidategroups $\Leftarrow$ all the groups;
    While *Candidategroups* $\neq \oslash$ **do**
      $C_g \Leftarrow$ Select a group from *Candidategroups* according to group pheromone trail;
      Candidates $\Leftarrow \{s_{ij} \in C_g$ that do not violate resource constraints in (6)-(9);
      Update local heuristic values by equation (12);
      Choose an object $s_{ij} \in Candidates$ with probability computed by equation (10);
      $S_k \Leftarrow S_k \cup s_{ij}$; Remove C_g from *Candidategroups*;
    Endwhile
    If $F(S_k) > F(S_{internalbest})$ **then** $S_{internalbest} \Leftarrow S_k$;
  Endfor
  If $F(S_{globalbest}) < F(S_{internalbest})$ **then** $S_{globalbest} \Leftarrow S_{internalbest}$;
    Update database and pheromone trails lower and upper bounds by τ_{min} and τ_{max};
Until maximum number of cycles reached (N_{max}) or optimal solution found;

The comparison time complexity of our algorithm with the recently proposed related algorithms are shown in Table 2.

Table 2. The comparison time complexity of the difference algorithms considered

Algorithm	Time complexity
Integer linear programming	$O(2^{MN})$
Exhaustive search	$O(M^N)$
Monte Carlo	M^q (where q is an integer smaller than N)
DiGA	$O(iter_{max} \bullet P_{max} \bullet W_{max} \bullet (N + log(P_{max} \bullet W_{max}))$
PAGA	$O(iter_{max} \bullet N \bullet L + iter_{max} \bullet L \bullet logL)$
MMAS	$O(N_{ant} \bullet N_{max} \bullet N \bullet M \bullet k)$

5 Experiments and Results

5.1 Dataset and Parameter Implementation

There are three factors that determine the size of the QASS problem: the number of tasks in composite service N, the number of candidate services per class M and the number of QoS constraints k ($1 \leq k \leq 4$). In the experiment, we present an experimental evaluation of our algorithm, focusing on its efficiency in terms of execution time and the quality in terms of the obtained best solution fitness value, and compare it with the recently proposed related algorithms Exhaustive search [8], Monte Carlo, DiGA [11], PAGA [13], SPSO [15], MPDACO [16],and CASS [17] on different scale test instances. We conducted experiments using the

QWS real dataset which includes measurements of 9 QoS attributes for 2507 real web services [23]. These services were collected from public sources on the web, including UDDI registries, search engines, and service portals, and their QoS values were measured using commercial benchmark tools [7]. The different scale test instances are created and shown as in Table 3.

Table 3. Main features of test instances

Instances	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15
N	5	5	5	5	5	5	5	5	5	5	10	10	10	10	10
M	4	8	12	16	20	24	28	32	36	40	50	100	150	200	250
w_k	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25

All algorithms are implemented in C language and executed on a Chipset Duo-core 3.0GHZ, 4GB RAM computer. The colony size of our algorithm is set as $N_{Ants} = 50$, and the other parameters are set as follows: $k = 10$, $\alpha = 1$, $\beta = 5$, $P = 0.01$, $\tau_{min} = 0.01$, $\tau_{max} = 8$, $\rho = 0.5$, $N_{Max} = 500$. The parameters of the other three compared algorithms are set as the same as the original papers [8],[11],[13],[15],[16],[17].

5.2 Case Study 1

In the first experiment, we implemented and evaluated the execution time and the quality in terms of the obtained best solution fitness value of the our algorithm and compare it with Exhaustive search [8], Monte Carlo, DiGA [11], PAGA [13], SPSO [15], MPDACO [16], CASS [17] algorithms on instances #1 to #10. In Fig.2(a) shows the optimality of algorithms with different number of candidate services change from 4 to 40. The ratio between the optimal QoS utility value obtained by the given approach and the optimal QoS utility value obtained by Exhaustive search approach are shown in Fig.2(b).

The results show that our algorithm is able to achieve above 98% optimality, which outperforms other approaches in Fig.2(a). Therefore, from Fig.2(b), we can s that MMAS algorithm is faster than the other three compared algorithms for the test instances.

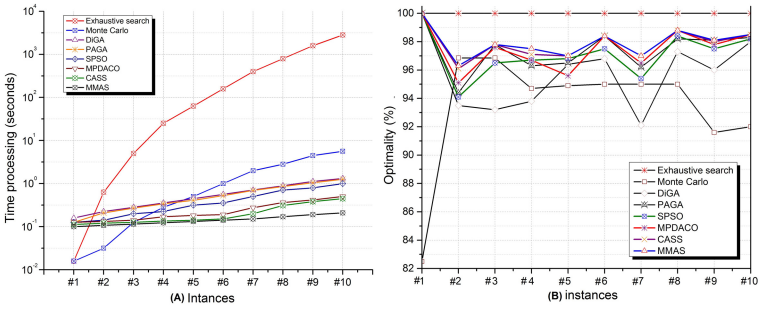


Fig. 2. Comparison time processing and optimality of algorithms considered

5.3 Case Study 2

In the second experiment, we run each DiGA, PAGA, SPSO, MPDACO, CASS and MMAS algorithms twenty times on each test instance #11 to #15. The termination condition for all the algorithms on each test instance is set based on the maximum candidate evaluation number, which is set as 3×10^4. The obtained best, worst, and average fitness values are given in Table 4 below. In which, the optimal value in the best, worst, and average case is shown in bold. The experimental results show that the MMAS algorithm achieved good quality and stable solution than the previous approach.

Table 4. Comparison of the results obtained by the difference algorithms considered

	DiGA			**PAGA**			**SPSO**		
Instance	*Best*	*Worst*	*Average*	*Best*	*Worst*	*Average*	*Best*	*Worst*	*Average*
#11	**0.1219**	0.1239	0.1231	**0.1219**	0.1256	0.1229	**0.1219**	0.1241	0.1228
#12	0.1321	0.1365	0.1345	0.1319	0.1351	0.1342	0.1316	0.1347	0.1339
#13	**0.1455**	0.1614	0.1555	0.1459	0.1582	0.1548	0.1468	0.1573	0.1536
#14	0.1595	0.1798	0.1677	0.1563	0.1611	0.1594	0.1546	0.1561	0.1549
#15	0.1474	0.1637	0.1562	0.1427	0.1532	0.1482	0.1469	0.1519	0.1476
	MPDACO			**CASS**			**MMAS**		
Instance	*Best*	*Worst*	*Average*	*Best*	*Worst*	*Average*	*Best*	*Worst*	*Average*
#11	**0.1219**	0.1237	0.1227	**0.1219**	0.1234	0.1225	**0.1219**	**0.1231**	**0.1223**
#12	0.1316	0.1342	0.1335	**0.1312**	**0.1337**	0.1332	**0.1312**	**0.1337**	**0.1331**
#13	**0.1455**	0.1501	0.1481	**0.1455**	0.1529	0.1481	**0.1455**	**0.1497**	**0.1473**
#14	**0.1521**	0.1562	0.1539	**0.1521**	0.1559	0.1533	**0.1521**	**0.1559**	**0.1531**
#15	0.1426	0.1503	0.1464	**0.1415**	**0.147**	0.1447	**0.1415**	**0.147**	**0.1446**

6 Conclusion and Future Works

In this paper, we study the E2E QoS constraint issue for composite business processes that are built using the Web service framework. We have presented and analysised service selection algorithms. The objective of the algorithms is to maximize user-defined service utilities while meeting the E2E performance constraint. We propose a MMAS algorithm to solve the problem. We evaluated our algorithm, focusing on its efficiency in terms of execution time and the quality in terms of the obtained best solution fitness value, and compare it with the recently proposed related algorithms Exhaustive search [8], Monte Carlo, DiGA [11], PAGA [13], SPSO [15], MPDACO [16], and CASS [17] on different scale test instances. The computational results showed that our approach is currently among the best performing algorithms for this problem. Service selection algorithms under dynamic QoS dataset is our next research goal.

References

1. Yu, T., Lin, K.J.: Service selection algorithms for Web services with end-to-end QoS constraints. In: Proceeding of ISeB, pp. 103–126. Springer (2005)
2. Ma, X., Dong, B.: AND/OR Tree-based Approach for Web Service Composition. Journal of Computational Information Systems 5(2), 727–728 (2009)
3. Yu, T., Zhang, Y., Lin, K.J.: Efficient Algorithms for Web Services Selection with End-to-End QoS Constraints. ACM Trans. on the Web 1, 1–26 (2007)

4. Zeng, L., Benatallah, B., Dumas, M., Kalagnanam, J., Sheng, Q.Z.: Quality Driven Web Service Composition. In: Proc. 12th International WWW Conference (2003)
5. Shankar, M., De Miguel, M., Liu, J.W.S.: An end-to-end QoS management architecture. In: Proceeding of 5th Real-Time Technology and Applications Symposium (1999)
6. Michlmayr, A., et al.: E2E support for QoS-Aware service selection, binding, and mediation in VRESCo. IEEE Trans. on Services Computing 3(3), 193–205 (2010)
7. Masri, E.A., et al.: Investigating web services on the world wide web. In: Proceedings of the International Conference on World Wide Web, China, pp. 795–804 (2008)
8. Ardagna, D., Pernici, B.: Adaptive service composition in flexible processes. IEEE Trans. on Software Engineering 33(6), 369–384 (2007)
9. Zeng, L., Benatallah, B.: QoS-Aware Middle-ware for Web Services Composition. IEEE Transactions on Software Engineering 30(5), 321–322 (2004)
10. Wu, M., et al.: Qos-driven Global Optimization Approach for Large-scale Web Services Composition. Journal of Computers 6(7), 1452–1456 (2011)
11. Zhang, C., et al.: DiGA: population diversity handling genetic algorithm for QoS-Aware web services selection. Computer Comm. 30(5), 1082–1090 (2007)
12. Liu, S.L., et al.: A Dynamic Web Services Selection Algorithm with QoS Global Optimal in Web Services Composition. Journal of Software 18(3), 651–655 (2007)
13. Su, K., Liangli, M., Xiaoming, G., Yufei, S.: An Efficient Parameter-adaptive Genetic Algorithm for Service Selection with End-to-end QoS Constraints. Journal of Computational Information Systems 10(2), 581–588 (2014)
14. Su, K., et al.: An Efficient Discrete Invasive Weed Optimization Algorithm for Web Services Selection. Journal of Software 9(3), 709–715 (2014)
15. Fan, X.Q., et al.: Research on Web service selection based on cooperative evolution. Expert Systems with Applications 38(8), 9736–9743 (2011)
16. Xia, Y.M., et al.: Optimizing services composition based on improved ant colony algorithm. Chinese Journal of Computers 35(2), 270–281 (2012)
17. Zhang, C., Yin, H., Zhang, B.: A Novel Ant Colony Optimization Algorithm for Large Scale QoS-Based Service Selection Problem. Discrete Dynamics in Nature and Society 2013, 1–9 (2013)
18. Liu, Y., Ngu, A.H.H., Zeng, L.: QoS Computation and Policing in Dynamic Web Service Selection. In: Proceedings of the International World Wide Web Conference, pp. 66–73 (2004)
19. Ran, S.: A Model for Web Services Discovery With QoS, pp. 6–9. ACM (2003)
20. Seekda, `http://webservices.seekda.com/`
21. Stutzle, et al.: Max-min ant system. Future Generation Computer Systems, 889–914 (2000)
22. Le, D.-N.: Evaluation of Pheromone Update in Min-Max Ant System Algorithm to Optimizing QoS for Multimedia Services in NGNs. In: Satapathy, S.C., Govardhan, A., Raju, K.S., Mandal, J.K. (eds.) Emerging ICT for Bridging the Future - Volume 2. AISC, vol. 338, pp. 9–18. Springer, Heidelberg (2014)
23. Al-Masri, E., Mahmoud, Q.H.: The qws dataset qmahmoud/qws/index.html, `http://www.uoguelph.ca/`
24. `http://randdataset.projects.postgresql.org/`

Artificial Intelligence and Machine Learning

Implementation Discrete Cosine Transform and Radial Basis Function Neural Network in Facial Image Recognition

Marprin H. Muchri, Samuel Lukas, and David Habsara Hareva

Computer Science Department, Universitas Pelita Harapan, Karawaci, Indonesia
Amrprin93@gmail.com, {samuel.lukas,david.hareva}@uph.edu

Abstract. Facial image recognition has been widely used and implemented in many aspects of life such as, investigation area or security purposes. However, research in this area is still been done. Source images of this paper are taken from image library provided in description of the Collection of Facial Images, 2008. This paper explains how 35 faces in JPG format with dimension 180 x 180 pixels can be represented by only 3 x 3 DCT coefficients and can be recognized fully 100% by Radial Basis Function Network.

Keywords: Discrete Cosine Transforms, Radial Basis Function, Image recognition.

1 Introduction

Two categories in face recognition system are feature-based approach and brightness-based approach. The feature-based approach is done by a special processing of face image that are key points of the face, such as the edges, eyes, noise, mouth or other special characteristics. The calculation process only covers some partial images that priory have been extracted in a certain way. On the other hand, the brightness-based approach calculates all parts of the image. Therefore it is also known as holistic-based approach or image-based approach.

Since all parts of the image have to be considered, the brightness-based approach needs a longer time to process and is also complicated. To make it short and simple, the image has to be transformed to a certain model. Many models have been proposed. One of that model was introduced by Turk and Pentland, 1991, using the Principle Component Analysis [1][2][3]. Another proposed model applied the Discrete Wavelet Transform (DWT) [4]. In the present paper, the DCT is applied.

After extraction of the features of the image, the system goes to a recognition system. There are also a lot of recognition models that can be used, such as back propagation neural network [4] and Hidden Markov Models [5][6]. This paper discusses on how face recognition can be done by using Radial Basis Function Network, RBFN.

R. Intan et al. (Eds): ICSIIT 2015, CCIS 516, pp. 113–118, 2015.
DOI: 10.1007/978-3-662-46742-8_10

2 Discrete Cosine Transform and Radial Basis Function Network

Discrete Cosine Transform (DCT) is a transform coding mostly used in signal processing or digital image processing. It is derived from the Discrete Fourier Transform (DFT). The objects of this paper are in the form of images. Therefore 2D-DCTs are implemented [7][8]. Special domain of an image $I(x, y)$ is transformed to frequencies domain as $C(u, v)$ is stated as (1) &(2) and the inverse as (3)

$$C(u,v) = \alpha(u)\alpha(v)\sum_{r=0}^{row}\sum_{c=0}^{col} I(r,c)\cos\left[\frac{(2r+1)u\pi}{2N}\right]\cos\left[\frac{(2c+1)v\pi}{2N}\right] \quad (1)$$

$$\alpha(u), \alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & for\ u, v = 0, \\ \sqrt{\frac{2}{N}} & for\ \ u, v = 1, 2, \dots, N-1 \end{cases} \quad (2)$$

$$I(r,c) = \sum_{u=0}^{row}\sum_{v=0}^{col} \alpha(u)\alpha(v)C(u,v)\cos\left[\frac{(2r+1)u\pi}{2N}\right]\cos\left[\frac{(2c+1)v\pi}{2N}\right] \quad (3)$$

DCT has important properties. They are de-correlation, energy compaction, domain scaling, separability, and symmetry. De-correlation means that there is no correlation in calculating among all the DCT coefficients. Therefore, all DCT coefficients can be calculated independently. DCT exhibits excellent energy compaction for highly correlated images. Efficacy of a transformation scheme can be directly gauged by its ability to pack input data into as few coefficients as possible without introducing visual distortion in the reconstructed image significantly. DCT is also not scaling invariant. This implies that in an image recognition system, all of the images that are used for training or identification have to be uniform. Separability means that the DCT coefficients can be computed in two steps by successive 1-D operations on rows and columns of an image. It is stated in (4). Another look at the row and column operations in Equation (4) reveals that these operations are functionally identical. Such a transformation is called a *symmetric transformation* [9].

$$C(u,v) = \alpha(u)\alpha(v)\sum_{r=0}^{row}\cos\left[\frac{(2r+1)u\pi}{2N}\right]\sum_{c=0}^{col} I(r,c)\cos\left[\frac{(2c+1)v\pi}{2N}\right] \quad (4)$$

The idea of Radial Basis Function (RBF) Networks derives from Multi-Layer Perceptron (MLP) networks but RBF Networks take a slightly different approach. They have five main features. They are two-layer feed-forward networks. The hidden nodes implement a set of radial basis functions (e.g. Gaussian functions). The output nodes implement linear summation functions as in an MLP. The network training is divided into two stages: first the weights from the input to hidden layer are determined, and then the weights from the hidden to output layer. The training/learning is very fast. Configuration of RBF Network for P input nodes with Q hidden nodes and R output nodes can be seen in [10]. Figure 1 is the scheme of RBFN.

The goal of RBF is to find a function $f: x^p \rightarrow y^r$ so that it can interpolate of a set of N data points in a p-dimensional input space, $X = (x_i\ x_2\ \dots\ x_p)$, to be mapped onto the r-dimensional output space, $Y = (y_i\ y_2\ \dots\ y_r)$. Radial Basis function of every hidden node has a center vector, $x_c = (x_{c_1}\ x_{c_2}\ \dots x_{c_p})$ and a variance, σ^2. The output of every

hidden node is stated in (5) then by doing linear combination with the weights, W_{kj}, from hidden nodes to the output nodes the the output of the RBF is performed (6).

$$\varphi(\|x - x_c\|),\ \varphi(a) = exp\left(-\frac{a^2}{2\sigma^2}\right) \tag{5}$$

$$y_k^* = f(x^*) = \sum_{j=1}^{q} W_{kj}\varphi_j\left(\left\|x^* - x_{c_j}\right\|\right) \tag{6}$$

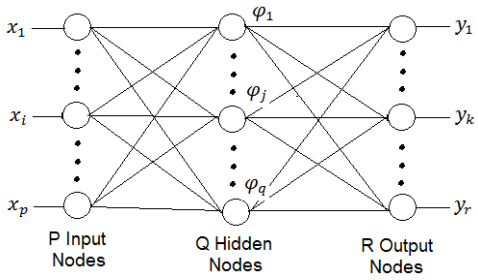


Fig. 1. Structure of RBFN

3 System Design

Block diagram of the system is presented in Fig. 2. They are training process and recognizing process. The output of training process is the weights of RBFN from hidden nodes to the output nodes, whereas for recognizing process is the name of the input facial image.

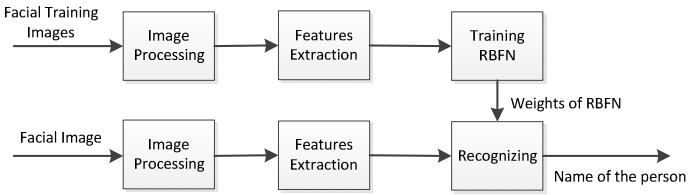


Fig. 2. Block diagram of the System

Facial training images consist of thirty five images with seven subjects. Each subject consists of five sample images. They are shown in Figure 3. Each image is normalized with the size of 64 x 64 pixels and transformed into gray scale image. By implementing DCT transform in feature extraction process than 64 x 64 DCT coefficients are performed. Some of these DCTs are trained in training process by RBFN.

Suppose p-DCT coefficients are chosen for the features of a facial image, then the input data set is a matrix, $X = \{x_{ij} \,|i = 0,2,\ldots,34\ ;\ j = 0,2,\ldots,p-1\}$ whereas the output training data set is a matrix, $T = \{t_{ik} \,|i = 1,2,\ldots,35\ ;\ k = 1,2,\ldots,7\}$. If image i in data set represents person number k in which $k = (i\ div\ 5) + 1$ then t_{ij} is defined in (7).

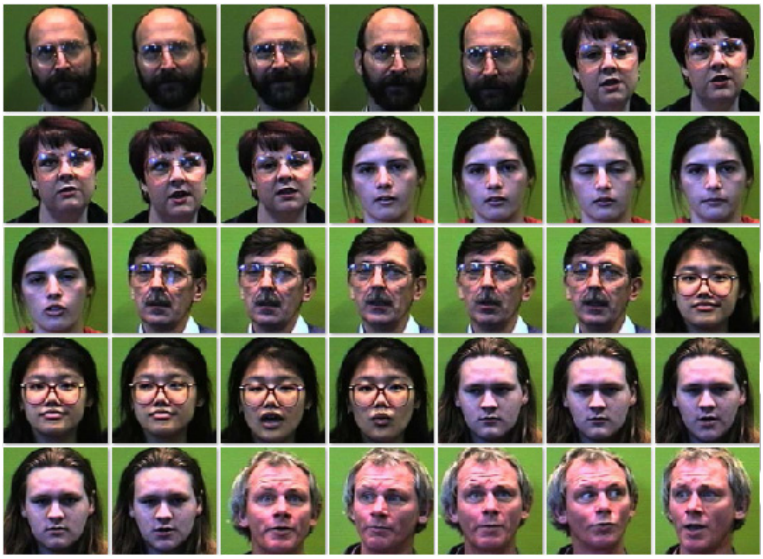


Fig. 3. Thirty five images to be trained

The output of hidden layer is formed in $H = \{h_{ik}\}$ in (9) after finding the center vector of hidden layer $C_k = \{c_{kj}, k = 1,2,\dots,7\,; j = 0,2,\dots,p-1\}$ and σ_k^2 in (8) and the weights is $W = \{w_{kk}\}$ in (10).

$$t_{ik} = \begin{cases} 1 & if\ k = (i\ div\ 5) + 1 \\ 0 & others \end{cases} \tag{7}$$

$$c_{kj} = \frac{1}{5}\sum_{i=5(k-1)}^{5k-1} x_{ij} \qquad k = \{1,\dots,7\}\ ;\ j = 0,1,\dots,p-1$$

$$\sigma_k^2 = \frac{1}{5p}\sum_{i=5(k-1)}^{5k-1}\sum_{j=0}^{p-1}(x_{i,j} - c_{kj})^2 \quad k = 1,2,\dots,7\,; j = 0,1,\dots,p-1 \tag{8}$$

$$h_{ik} = \varphi_k(x_{ij}) = e^{\left(\frac{\|x_{ij} - c_{kj}\|^2}{2\sigma_k^2}\right)} \qquad i = \{0,\dots,34\}, k = \{1,\dots,7\} \tag{9}$$

$$W = (H^t H)^{-1} H^t T \tag{10}$$

In recognizing process, a single facial image is represented by $x_j \quad j = 0,2,\dots,p$ then the output of recognizing process, $O = \{O_j \mid j = 1,2,\dots,7\}$ is in (11)

$$h_k = \varphi_k(x_{ij}) = e^{\left(\frac{\|x_j - c_k\|^2}{2\sigma_k^2}\right)} \qquad k = \{1,\dots,7\}$$
$$O = HW \tag{11}$$

It is clear that index of the highest value of component O_j indicates that input facial image belongs to that index number of person in the system.

4 Experiment Result and Discussion

The experiment is to determine how many DCTs coefficient is needed to achieve better recognition percentage. It is done from 2 x 2 DCTs to 4 x 4 DCTs. The results all of 35 new input images, figure 4, can be recognized 100% correct. The result is tabulated at Table 1.

Fig. 4. Thirty five images to be recognized

Table 1. The percentage and threshold results of the experiments

Subject	DCT 2 x 2		DCT 3 x 3		DCt 4 x 4	
	% rec	Threshold	% rec	Threshold	% rec	Threshold
1	100	0.60	100	0.72	100	0.65
2	100	0.80	100	0.77	100	0.74
3	100	0.51	100	0.62	100	0.59
4	100	0.81	100	0.93	100	0.94
5	100	0.90	100	0.96	100	0.96
6	100	0.84	100	0.93	100	0.92
7	100	0.96	100	1.00	100	0.96

From the system design, it indicates that from a single input facial image, it will be 7 values of each output. The values indicate the percentage recognition to the index of associated person. From the table, using DCTs 2 x 2 subject no-1, the threshold is 0.6. It means if the value of the first output value is at least 0.6, it indicated that the input facial image belongs to the first person. Whereas threshold value by using DCTs 3 x

3 of subject 6 is 0.93, it means the input image belongs to person number 6 if the sixth value of the output at least 0.93. From that table, the best DCTs is 3 x 3.

5 Conclusion

System is able to recognize 100% facial input data using Discrete Cosine Transform and Radial basis Function Network. However, some works still can be done if the facial data is not taken from the facial image database, such as from direct camera.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
2. Diamantaras, K.I., Kung, S.Y.: Principal Component Neural Networks: Theory and Applications. John Wiley & Sons, Inc. (1996)
3. Pentland, A., Moghaddam, B., Starner, T.: View-Based and Modular Eigenspaces for Face Recognition. In: IEEE Conf. on Computer Vision and Pattern Recognition, MIT Media Laboratory Tech. Report No. 245 (1994)
4. Alwakeel, M., Shaaban, Z.: Face Recognition Based on Haar Wavelet Transform and Principal Component Analysis via Levenberg-Marquardt Back propagation Neural Network. European Journal of Scientific Research 42(1), 25–31 (2010) ISSN 1450-216X
5. Kohir, V.V., Desai, U.B.: Face Recognition Using a DCT-HMM Approach. Indian Institute of Technology, Mumbai (1998)
6. Handy, S.E., Lukas, S., Margaretha, H.: Further Tests for Face Recognition Using Discrete Cosine Transform And Hidden Markov Model. In: Proceeding, International Conference on Electrical Engineering and Informatics (MICEEI), Makasar (2012)
7. Khayam, S.A.: The Discrete Cosine Transform (DCT): Theory and Application. Michigan State University (2003)
8. Acharya, T., Ray, A.K.: Image Processing: Principles and Applications. John Wiley (2005)
9. Radha, H.: Lecture Notes: ECE 802 - Information Theory and Coding (January 2003)
10. Oh, B.-J.: Face Recognition Using Radial Basis Function Network based on LDA. International Journal of Computer, Information Science and Engineering 1, 401–405 (2007)

Implementation of Artificial Intelligence with 3 Different Characters of AI Player on "Monopoly Deal" Computer Game

Irene A. Lazarusli, Samuel Lukas, and Patrick Widjaja

Department of Informatics, Faculty of Computer Science, Universitas Pelita Harapan
MH Thamrin Boulevard, Lippo Village, Karawaci, Tangerang
{irene.lazarusli,samuel.lukas}@uph.edu

Abstract. Monopoly Deal is a card game that derived directly from the traditional Monopoly Game. The game consists of a set of cards as a game device. This game aims to collect three different sets of property cards. This research has been conducted to explore the potential for further developing this simulation game, by implementing rules and strategies of Monopoly Deal. BFS method was used as the algorithm of the game. The character of players in the game is divided into three kinds, namely: Aggressive, Defensive, and Normal. Each character would make a different influence on the provision of assets and on the selection of card or money being played. The implementation of the game consists of characters, check assets system, and the computer-based card game itself with AI intelligence player. The testing results show that the combination of aggressive character and logical asset as the highest winning rate of 45%.

1 Introduction

Monopoly is a board game created by Parker Brothers on 5 November 1935. Its name was taken from an economic concept which means dominate the entire market [1]. In 2008, a card game named Monopoly Deal has been developed based on the conventional Monopoly game. Unlike the conventional one- that uses a board, a set of property cards, monopoly money sheets, dice, and some pawns as media to play, Monopoly Deal only uses a set of cards as the media.

At the conventional monopoly, some pawn is used as a representation of the player, while the Monopoly Deal doesn't used any object to represent the player, because in this game each player plays his cards on the game area / game table. This game is considered complete when a player has first collected three full sets of property cards. Monopoly Deal game is being played using 106 sheets of card consist of: 20 money cards, 34 action cards, 13 rent cards, 28 property cards, and 11 property wildcards[2], where each card has its own function and purpose. Each type of card can be used by players in certain circumstances based on the rules and strategies of the game.

Monopoly has being evolved in the world of video games, with various game themes. However, the Monopoly Deal has yet developed neither in a video or

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 119–127, 2015.
DOI: 10.1007/978-3-662-46742-8_11

computer game. Therefore this study was made to develop a computer game based on Monopoly Deal card game.

1.1 The Rule of Monopoly Deal

Monopoly Deal game can be played by two to five players. Each player is dealt five cards at the beginning of the game. The objective of each player in this game is to collect three sets of properties with 3 different colours. When a player gets a turn, he draws two cards from the deck and can play up to three cards per turn. A play is considered as any time a card is played on the table, such as money into your bank, property cards on the table, action cards played into the middle, etc. The player can use three kinds of cards when his turn took place, namely: property cards, money cards, and action cards. At the end of each player's turn, they cannot have more than 7 cards in their hand. If they do, they need to discard the excess Monopoly cards into the discard pile in the middle. The final goal of Monopoly Deal is being the first player who complete 3 full property sets on the table in front of the player [3]. There are three places where cards can be played during a turn:

1. A player can place money cards or action cards face up in their bank.
2. A player can play property cards face up in front of them in their property section.
3. A player can play action cards in the middle discard pile.

There are a total of 34 action cards include: rent card, house card, hotel card, force-deal card, pass-go card, etc. There are a total of 28 property cards in the Monopoly Deal, contains 10 different coloured property sets and include: Green Set, Dark Blue Set, Light Blue Set, Red Set, Utility Set, Yellow Set, Orange Set, Brown Set, Purple Set, and Railroad Set. Each property card tells the player how many cards are needed of that colour to complete the set. For example, the brown set only needs 2 brown property cards to be completed and the black railroad set needs 4 railroad property cards to be a completed set. Monopoly Deal game also has a few other rules, such as:

1. **Payment Rules**: the player who pay may determines the card used for payment, either using the money cards or property cards, and there is no change for every payment.
2. **Property Cards Usage Rules**: The multi-coloured-property wild cards can be used as any colour property card, but they don't have value of money. It is not allowed to construct more than one property card of the same colour in two sets of properties. The property card can be taken by another player using the Sly-Deal and Force-Deal cards.
3. **Action Cards Usage Rules**: The Double-the-Rent cards can be used to double the rental value. The Just-Say-No cards can be used to cancel/against the previous player's action card. A Just-Say-No card can be cancelled by another Just-Say-No card. A Just-Say-No card doesn't count as a card turn. The Hotel cards and House card can't be played on the Railroad card and Utility card. The Pass-Go card can be used to pick up two cards from the deck. The "It's-My-

Birthday" card and Debt-Collector card can be used to ask for money from other players.

1.2 The Strategy of Monopoly Deal

Monopoly Deal has a wide range of strategies to win the game, such as money strategy, property management strategy, the strategy of using the action cards, or a strategy for investment in the game. There are some of the strategies in Monopoly Deal, namely: Money Strategy, Memory Strategy, Protection Strategy, Force-Deal Strategy and Sly-Deal Strategy, Investment Strategy, and Deal Breaker Strategy [4]. The money strategy emphasizes collecting money in the other player's bank. Memory strategy is associated with recalling the card had been used. Protection strategies is related to protecting property cards owned by the player cards and increase the protection while the opponents charge the player with bank regulation, property, and payments. Sly-Deal and Force-Deal strategy is used when the player will take the property of the opponent's card. This strategy is used when the player will take a complete set of property cards belong to the opponent. In addition to the strategies mentioned above, there are also a few strategies that are not implemented in this discussion, among others: Card Layout Strategy, Strategy Mind Games, and Strategy Player Weakness.

2 Research Method

This section explains about the calculation and division of game phase, and the implementation of various strategies such as card elimination, card arrangement and game asset management. It also explains about the implementation of BFS method during each phase.

2.1 Game Phase

The game is divided into three phases. The first phase refers to the earlier rounds of the game. The second phase showed that the player was already in the middle of game time. The third one was meant to be almost at the end of the game and keep on in this phase, although the pile of cards already re-shuffled. The calculation of game phase depends on the game cycle. One cycle means that each player has had one turn to play. The number of game cycle can be calculated by:

$$\text{Number of cycles} = \frac{(K - (B \times N) - P)}{D \times N} \tag{1}$$

N = number of players
K = number of cards
B = card being dealt in the beginning
P = *Pass-Go* card being used
D = card being dealt on every turns

2.2 Card Elimination

When a player has more than seven cards at the end of the turn, he has to eliminate one of his cards. The implementation phases by computer player on this event are as follows:

1. Check the value of entire card
2. Choose the card with the lowest value for being eliminated
3. If there are at least two cards got the same value, the player will eliminate the card in accordance with the value of money on the card. For example, the lowest value is on the *Double-the-rent card* because the absence of *Rent cards* for being used together. Another card with the low value is the *House card* that wasn't yet belongs to any complete set of properties. Then the player should choose to eliminate the *Double-the-rent card* with value of M1 rather than the *House* with value of M3.

2.3 Card Arrangement

The BFS method will also applied when the computer do arrangement on the hand card. Example of this arrangement is in the case when computer player- who already has red and yellow *property card* on its hand - got the red and yellow colored *Property-Wild card* from another player. The decision about which color of card will be choose by computer in this example, are shown in Table 1.

Table 1. Decision for Red or Yellow Property-Wild Card

The number of red *property card*	The number of yellow *property card*	Decision	What color
0	0	Take the card	Yellow
0	1	Take the card	Yellow
0	2	Take the card	Yellow
0	3	Take the card	Red
1	0	Take the card	Red
1	1	Take the card	Yellow
1	2	Take the card	Yellow
1	3	Take the card	Red
2	0	Take the card	Red
2	1	Take the card	Red
2	2	Take the card	Yellow
2	3	Take the card	Red
3	0	Take the card	Yellow
3	1	Take the card	Yellow
3	2	Take the card	Yellow
3	3	Don't take the card	-

2.4 Implementation of Game Strategy

The strategy being applied in the game is Money Strategy, Protection Strategy, Force-Deal Strategy and Sly-Deal Strategy, and Deal Breaker Strategy. The implementation of the strategy described in previous section (1.2) is shown by the case example when the computer charge another player by using two-colored Rent card. When checking two-colored Rent card, computer will examine the value of each color of the Rent card. The sequence of card checking is done on the following step:

- Check out the first color of the Rent card.
- Check out the pile of property and get the value of property card with the first color.
- Check out the second color of the Rent card.
- Check out the pile of property and get the value of property card with the second color.
- Choose the rent value based on the color of property with the highest rent value.
- Check out the pile of money and the pile of property owned by another three players.
- If another player has a lower amount of money than the rent value of computer's rent card, then computer will check out the other player's property card. Then the players add up the value of money and property gained from the other players to be the value for decision-making to choose a card to play.
- If another player has a higher amount of money than the rent value of computer's rent card, then the computer will check the highest value of money card owned by the other players.
- If the computer player has the possibility of getting money or property which is three times higher than the rental value of the Rent cards, then the total of the rental value will be used as decision-making in choosing a card to play.

2.5 AI Player Characters and the Type of Game Asset

The characters of player are divided into three kinds, namely: aggressive, defensive, and normal. Each character has an influence on the valuating of assets and playing property cards or money. The game assets refer to money cards and property cards. These assets are used directly as computer examining the money cards and property cards in its hand. However, when the computer will play an action card, then the value of these assets will be used by examining the possibility of the increase of assets if the card is played. The game asset consists of three kinds of assets, namely: logical asset, value asset, and logical-value asset.

2.6 Breadth First Search Implementation

Implementation of Breadth First Search (BFS) was first performed when the computer (AI player) got turn to examine its own cards, both on hand as well as on table. The cards will be analyzed three times. When checking the *money cards* in its hand,

computer will check the value of the *money cards* with the total number of all cards on its pile of money in bank. When checking *property cards* in its hand, computer will check the color of the card with the total number of all *property cards* on its *property card* piles. On the *action card*, the computer will check the value of each *action card* based on the function of the card.

For example, computer owns six sheets of card, comprising: *money card* worth ₥1, *money card* worth ₥2, brown colored *property card*, yellow colored *property card*, pink and orange colored *rent card*, and debt collectors card. The BFS implementation is shown in figure 1 and figure 2.

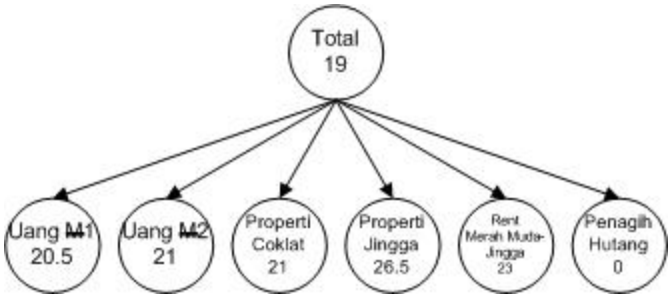
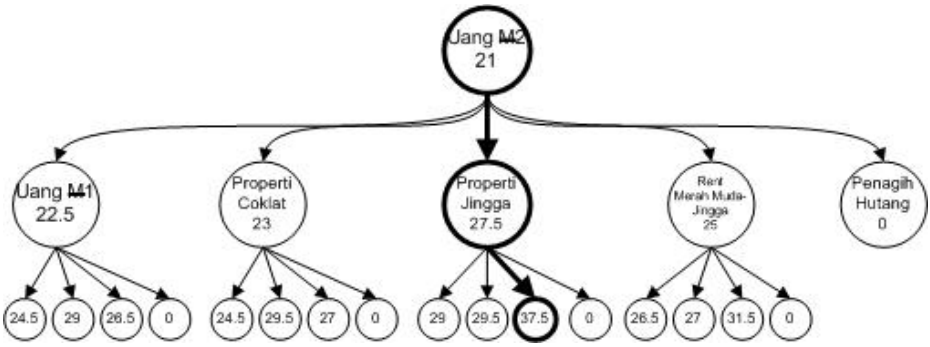


Fig. 1. BFS Tree Level 1



Search tree on the node "₥2 *money card*"

Fig. 2. Search tree on the node "₥2 *money card*"

3 Results and Analysis

3.1 Game Implementation

The design of Monopoly Deal card has a different variation from the original card design, but still based on the original one. Figure 3 shows the card design. The design of main menu of the game is shown on the figure 4(a). The regions of main playing table divided into the deck table, property table, the card on the hands, and the navigation panel, as shown in figure 4(b). The table located on all four sides of each

player's screen is used to put the on-hand player cards and the money cards being played. In front of each player, there are table used to lay the property cards being played by every player. On the right side of the screen, there is a navigation bar that can be used by players to return to the main menu, look at the types of cards, see the short help, see the history of card being played by each player, the "pass" button to pass the player's turn, the icon of the player who is in turn, and the number of cards that can be played during the player's turn.

Fig. 3. The design of cards used in Monopoly Deal Game.

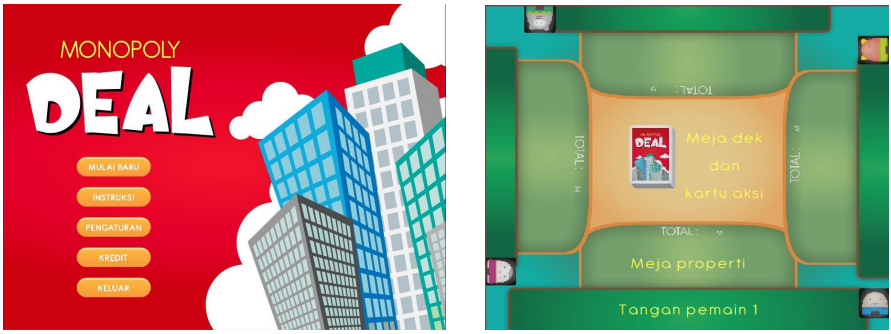


Fig. 4. The design of main menu and main table of Monopoly Deal Game

3.2 Game Testing

Studies continue to examine every character of the player with the three kinds of game assets. The test is performed in four stages, namely:

Stage#1: All computer players use the *Aggressive* character, with computer player#1 uses *logical asset*, computer player#2 uses *value asset*, and computer

player#3 uses average *logical-value asset*. The results of this testing showed that *aggressive* character will work more effectively when using *logical* assets.

Stage#2: All computer players use the *Defensive* character and each player using *logical assets*, *value assets*, and the average *logical-value assets*, respectively. In this scenario, the player#2 with combination of *defensive* character and *value asset* got the highest winning rate.

Stage#3: All computer players use the *Normal* character with the combination of *logical asset* on computer player#1, *value asset* on player#2, and average *logical-value asset* on player#3. The highest winning rate is obtained by the player#1 with combination of *normal* character and *logical assets*.

Stage#4: The last test is performed by combining the best results of the three previous tests. Player#1 use the combination *aggressive* characters with *logical asset*, the combination of *defensive* character with the value asset on player#2, the *normal* character with *logical* are combined and applied on player#3. The combination of *aggressive* character and *logical* asset got the winning percentage of 45%.

From the game tests with the various characters of players – which had been held 20 times for each scenario of stage#1, #2, and #3, the result is shown on the table 2.

Table 2. Testing Results for 3 Characters of AI Player

Stage#	Character	Number of Game Winning by Player with			
		Human Player	*Logical* Asset	*Value* Asset	*Logical-Value* Asset
1	***Aggressive***	4	**7**	4	5
2	***Defensive***	3	5	**7**	5
3	***Normal***	6	**8**	5	1

After the previous three stages, in the stage#4 another 20 times of test is conducted to the game for 3 combination of character with particular types of asset. The result is shown on the table 3.

Table 3. Testing Results for the Best-3 Combinations of Character (Stage#4)

Character	Number of Game Winning by Player with			
	Human Player	*Aggressive-Logical Asset*	*Defensive-Value Asset*	*Normal-Logical Asset*
Total	4	**9**	5	2
Percentage	20%	**45%**	25%	10%

4 Conslusion

Monopoly Deal game application has been successfully implemented with Breadth First Search algorithm. Game strategy of Monopoly Deal, which is manifested into 3

different characters, has been applied to the application. The *aggressive* character of player has the highest winning rate in combination with the *logical asset* percentage of 35%. While the player equipped with *defensive* character got the highest winning rate in combination with *value asset* with a percentage of 35%. Highest winning rate achieved by the player with the *normal* character has being obtained in combination with *logical asset* with a percentage of 40%. The tests of three different character showed that the highest winning rate was 45%, occurred on the combination of *aggressive* character and *logical* asset.

References

1. The Great Idea Finder: Monopoly History – Invention of Monopoly internet, http://www.ideafinder.com/history/inventions/monopoly.htm (accessed December 22, 2012)
2. Monopoly Deal Strategy: Monopoly Deal Count & Types of Cards, http://monopolydealrules.com/index.php?page=cards#top (accessed September 23, 2012)
3. Monopoly Deal Strategy: Monopoly Deal: How to Play & How to Win, http://monopolydealrules.com/index.php?page=play#top (accessed September 23, 2012)
4. Monopoly Deal Strategy: Winning Monopoly Deal Strategies for Making Money in Monopoly Deal, http://monopolydealstrategy.com/index.php?page=money#top (accessed September 23, 2012)
5. Princeton University, Online Homepage: "Breadth-First Search", http://www.princeton.edu/~achaney/tmve/wiki100k/docs/Breadth-first_search.html (accessed October 13, 2012)

Optimizing Instruction for Learning Computer Programming – A Novel Approach

Muhammed Yousoof and Mohd Sapiyan

Department of MIS, Department of Computer Science
Dhofar University, Gulf University of Science and Technology
Salalah, Kuwait
Oman, Kuwait
m_yousoof@du.edu.om, sapiyan@gust.edu.kw

Abstract. Computer programming is a highly cognitive skill. It requires mastery of many domains. But in reality many learners are not able to cope with the mental demands required in learning programming. Thus it leads to rote learning and memorization. There are many reasons for this situation. However one of the main reasons is the nature of the novice learners who experience high cognitive load while learning programming. Given the fact that the novice learners lack well defined schema and the limitation of the working memory, the students could not assimilate the knowledge required for learning. Many types of learning supports like visualization is provide to reduce the cognitive load. It is expected that the visualization will help in reducing the cognitive load by expanding the working memory. However the effect of visualization in learning is not clearly tangible. There are two common methods used to measure the cognitive load namely physiological and non physiological measures .It is found based on our prior studies that non physiological measure are more appropriate for measuring cognitive load in a class room situation. The non physiological measures include NASA TLX rating scale. It is also observed from our prior studies that there is a variation in learning performance using same visualization support among the students in a homogenous group. This variation is attributed to the level of Long Term Memory which includes prior mathematical, prior language skills, demographics and gender. This paper will propose a framework to optimize the instruction to learners based on their background profile and the extent of the long term memory and will employ neural network to optimize the instruction by suggesting the best visualization tool for each learner. The paper also validates the performance of the proposed tool by a study with the learners to evaluate the accuracy of the tool.

Keywords: Programming, Visualization, Cognitive Load, NASA TLX, Neural Network Framework.

1 Introduction

Lot of research has been done on finding out the reasons and solutions to overcome the learning difficulties of computer programming .There are many prior studies done

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 128–139, 2015.
DOI: 10.1007/978-3-662-46742-8_12

on the difficulties in learning computer programming.[2][6][8][10][13][16][20][29][30][31][33][35].One of the main difficulty in learning programming is the Cognitive Load. Cognitive Load experienced during the learning of programming is high as the novice learners lack well built schema in the long term memory [22][32]. This problem could be addressed by using the visualization [12][21][24][25]as the working memory could be extended by using the visual aids [5]. However it is not clear how this visualization helps in managing the cognitive load. There are two ways to measure the cognitive load namely physiological measures [7][17][36] and non physiological measures [14][15]. Physiological measures are not so suitable in a class room environment based on our previous studies using EEG and GSR.The validity of the non physiological measures was done by another study using NASA TLX rating scale and in this study it was also found that the visualization help in learning varies also based on the other factors such as prior English knowledge, mathematics knowledge, demographics etc. It was also observed that the cognitive load also varies from subject to subject in learning the same concept using the same visualization tool and programming languages in a same homogenous group. This variation of the load is due to many factors which include demographics of the student and the schemata of the Long Term Memory which is based on the prior knowledge of the English language, Mathematical background and programming knowledge. It is difficult to determine the best tool for learning for every individual student as it depends on various factors mentioned before. So it becomes necessary to devise a suitable mechanism to predict the best tool for learning for each student considering many factors that affect learning.

Any problem whose output is associated with many factors will be easily represented using the Neural Network. There are many prior works available in the literature where neural network is used in real time examples such as sales price predication, stock price prediction [27], forecasting financial economical series [18] and real estate price prediction [34]. Artificial Neural networks have been used in many applications related to education fields like Intelligent Tutoring System. Artificial Neural network is highly successful in arriving at predictions where there is a high chance of uncertainty. It is a valid point that in our study higher element of uncertainty exists, as the cognitive load is complex and varies according to learning levels and ability of the learner.

2 Model for Optimizing Instruction

We decided to use the supervised learning method in our neural network model to determine the best visualization tool as an output. The network can be trained using input of cognitive load from the NASA TLX scale and also the Long Term Memory (LTM) which is based on the prior knowledge of Mathematics, English and programming. The weights for each of these inputs are simplified and discussed in the subsequent sections.

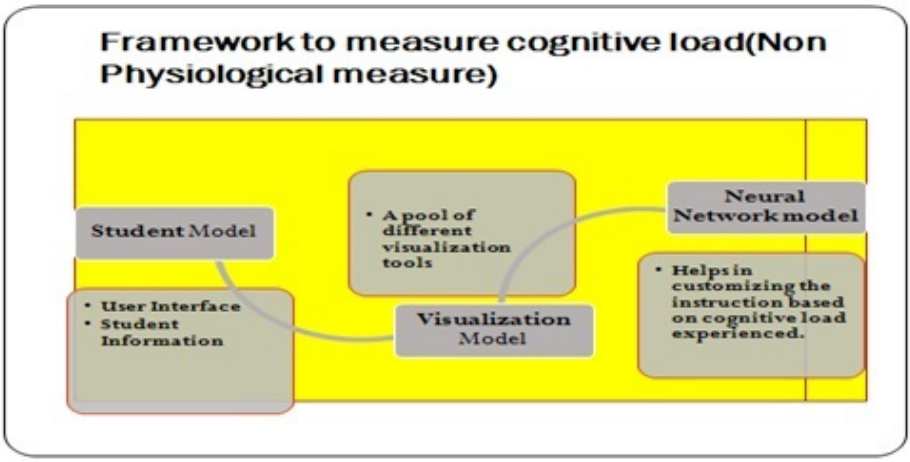


Fig. 1. Proposed Model of Optimizing Instruction for Learning

The above diagram shows the framework of the tool for optimizing learning is built. But at the present the integration of the visualization model with the neural network model is not accomplished. Both the models work as separate entities. Prediction of the tool is done by the neural network model and on the basis of the recommendation the appropriate visualization tool is recommended to the learners for learning the concepts.

3 Neural Network Model for Computer Program Learning

The Neural Network (NN) model based is composed of different layers. The input layer parameters include the various cognitive factors calculated using NASA TLX scale, Mathematical background, analytical background and test performance .The output layer of the network includes the recommendation of the tool for optimized learning. The predictions are based on the cognitive load of the learners as well as their performance in a particular task. The input to the neural network is chosen so that it accounts for both the working memory and long term memory. These two memories play an important role in learning as per the Cognitive Load Theory (CLT). The major inputs given for the neural network model are based on the prior knowledge of mathematics and computer programming, prior English knowledge and in addition to the load expressed using NASA TLX scale. The first two inputs represent the Long Term Memory (LTM) and the last input represents the working memory. The inputs to the neural network are simplified which will be discussed in the subsequent sections. This system has adopted the feed forward loop whereby successive iterations make the system more efficient in predictions. The network is represented in Figure 2

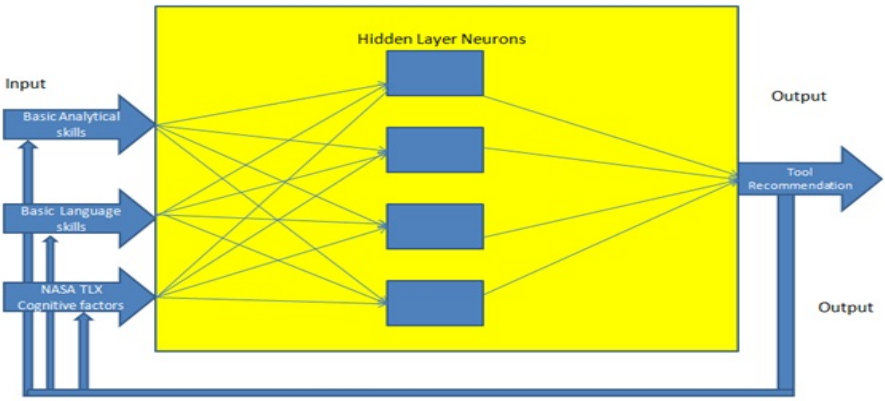


Fig. 2. Neural Network Model for Visualization selection

3.1 Long Term Memory (LTM) Schemata Calculation

Among the inputs for the neural network are value associated with the LTM, the language weight and the analytical weight of the LTM. The language weight includes the grades secured in the Pre University English, Matriculation English and IELTS test score. The analytical weight includes the grades of Matriculation Mathematics, Pre University Mathematics and weight-age of Pre University Computer Science grades. The input for the language weight is calculated by assigning 50% weight for the English Language skill test which was administered to them prior to learning programming, 25% weight-age for the English language grade in Matriculation and another 25% is considered from the Pre University English language grade. The sum of these weights is one hundred. It represents the input to a maximum of 1 which is the case of normal input to any neural network. We have assigned more weight for the IELTS type examination as it reflects the current state of the student's English knowledge. The maximum grades for English language skills test is 10 Points. The maximum marks for matriculation English and Pre University English are 100 and 200 respectively. The English language is 50 % of the load of the LTM. The language weight of the LTM is calculated using the formula given below.

$$\mathbf{LWLTM} \quad = \frac{\left(\frac{a}{2}\right)+\left(\frac{b}{4}\right)+\left(\frac{c}{8}\right)}{2} \tag{1}$$

where

a = IELTS marks
b = Matriculation English marks.
c = Pre University English marks.

The analytical weight of the LTM is also calculated in the same manner as the language weight by considering the Matriculation Mathematical marks Pre University

Mathematical grade and Computer Science marks in the weight-age as mentioned. The 50 % of weight is assigned to the matriculation mathematics marks remaining 50% is assigned equally to the analytical skill which is based on the Pre University Mathematics marks and Pre University Computer Science marks. The maximum marks for Matriculation Mathematics is 100 and Pre University Mathematics and computer science is 200.We have assigned more weight for Matriculation Mathematics marks as this mathematics forms the fundamental knowledge related to Mathematical concepts. Analytical weight of Long Term Memory is calculated as shown in the following formula.

$$\mathbf{AWLTM} = \frac{\left(\frac{d}{2}\right)+\left(\frac{e}{8}\right)+\left(\frac{f}{8}\right)}{2} \tag{2}$$

where

d = Matriculation Mathematics marks.
e = Pre university Mathematics marks.
f = Pre university Computer science marks.

3.2 Cognitive Load Calculation

Another input to the neural network is the cognitive load experienced during the task. The training data for the inputs of this neural network is based on the experimental data done in chapter 6.The cognitive load has two dimensions based on NASA TLX scale namely demands imposed on the subject (Mental, Physical and Temporal Demands) and the reaction of the subject with the task (Effort, Frustration and Performance).The demand imposed on the subject has to be minimal and the reaction of the subject must be positive (low scores) for the effective learning processes. The weight is the value determined based on the frequency of the particular attribute of the cognitive load reported to be challenging using the standard procedure of flip book. There are 15 possible combinations available in the flip book. Every occurrence of the particular attribute is rated as one. The total weight is the sum of occurrences of that particular aspect of cognitive load. The total weight is converted to percentage as shown below in the formula.

$$\mathbf{WCLI} = \frac{OW}{15} * 100 \tag{3}$$

Where

WCLI = Weight of cognitive load item
OW = Frequency of the cognitive load item

The rating scale for each of the six aspect of the cognitive load is measured using the scale of 20 in NASA TLX. The rating scale value is converted to percentage by the following formula.

$$RP = \frac{OR}{20} * 100 \quad (4)$$

Where

RP = Rating percentage
OR = Observed rating

4 Evaluation of the Proposed Framework

The Neural Network (NN) model had been tested by using a sample of 36 students. The subjects were in first year introductory programming course of the Bachelor degree program at a college in Chennai, India. The preliminary information of these subjects like demographic, Mathematical skills and English skills were collected. On the basis of the preliminary information collected the Neural Network tool was used to select the most appropriate visualization tool for learning programming. The neural network tool selected the tool based on the training data on a similar study using NASA TLX scale. The neural network model selected the appropriate tool from the pool of visualization tool and in this study and the two tools namely ViLLE which is referred as tool 1 and Teaching Machine which is referred as tool 2 in this paper were available in the pool of visualization tools.

The subjects learnt the two concepts of C++, namely classes and inheritance with the allocation of four hours to learn each concept. The study is done in such a way that the two concepts were learnt consecutively one after the other using the visualization tool chosen by the Neural Network and subsequently the students were made to learn the same concept in C++ namely classes and inheritance using the other tool. At the end of the learning, learning performance and cognitive load are measured. Learning performance was observed by the grades obtained in the test and cognitive load is measured by using NASA TLX scale. The 36 students comprised 18 boys and 18 girls. The sample represented all the three demographic areas namely rural, urban and semi urban. This is necessary for the validity of the system as the system uses the demographic area as one of the parameters for selecting the tool. The learners were distributed as in demographics as 33% of the learners from Urban,36% of the learners from Semi urban and 31% are from rural area from Rural.

It is noted that the Tool 2 visualization tool was selected as the most appropriate tool in most cases. The selection is based on the training data based on the previous study based on non physiological measures. Teaching machine was selected approximately 70 % during the training phase in the neural network tool. The Neural Network model selected the Tool 2 for 25 out of 36 students which accounts for approximately seventy percent of the subjects who took part in the system evaluation. This selection is justifiable as the students involved in the evaluation are a mix of various demographic factors like rural, urban and semi urban and also with mixed levels of English and mathematics background. The subjects were not informed about the selection of the appropriate visualization tool made by the neural network tool.

The results were first analyzed the results for learning performance the concepts using the two tools by considering the average performance of all learners. The results are shown in the following Fig 3

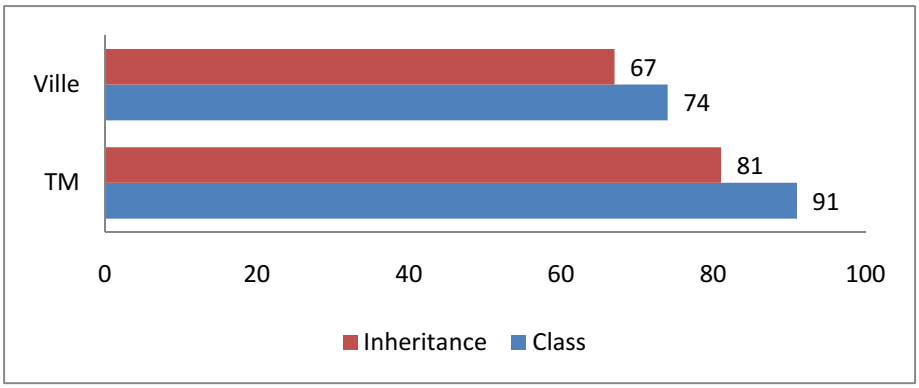


Fig. 3. Learners' performance using the selection made by the model

We can observe from the Figure 3 that the tool 2 has higher performance for the two concepts than tool1.

The performance while learning Classes was better with tool 2. The performance while learning inheritance was higher while using tool 2. So, on the basis of the higher performance of the Teaching Machine tool on an average justifies the selection of the Neural Network model, where 25 out of 36 students have been selected to learn with tool 2.

Another variable to determine the effectiveness of the tool is through the cognitive load experienced. The cognitive load is less while using the tool 2 but high while using the tool1. The following Figure 4 below shows the average of the cognitive load for all the subjects while learning both the concepts using the two visualization tools.

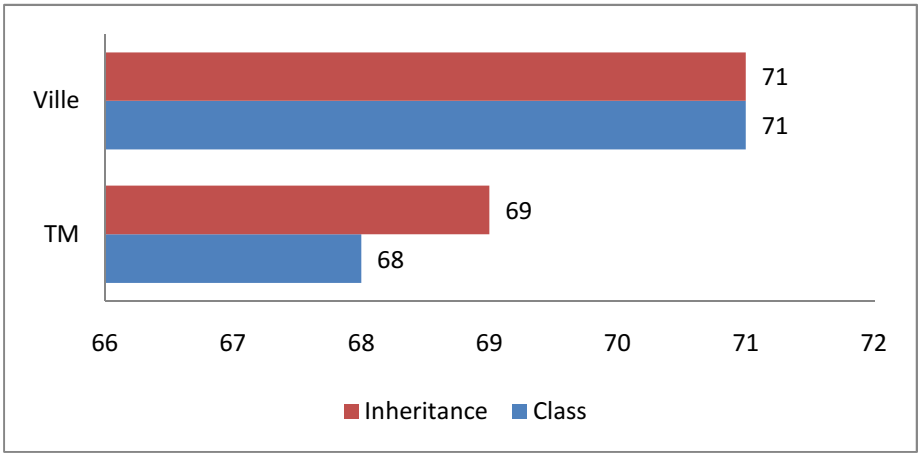


Fig. 4. Cognitive Load with Different Visualization Tools

The cognitive load was lower while using tool 2 in learning classes.

It is also noted that the cognitive load was lower while using tool 2 in learning inheritance. So on the basis of the lower cognitive load of the tool 2 on an average justifies the selection of the Neural Network model, where 25 out of 36 students were selected to learn with tool 2.The data is analyzed in different horizons according to the gender, demographic area and accuracy of the Neural Network system in selecting the best visualization tool. The results based on the analysis of gender is tabulated in the Table 1 as shown below

Table 1. Genderwise comparision of peformance for different visualizaiton tools

	Tool 2		**Tool 1**	
	Girls	Boys	Girls	Boys
Performance	85	85	70	72.5
Cognitive Load	67	69.5	69.5	73

The result summary of the students' performance and cognitive load based on gender is as given below. The performance of girls and boys while using tool2 was almost same. However the performance of the boys was higher while using Tool 1. Both boys and girls had lesser cognitive load while using Tool 2. The performance of the boys and girls was higher while using Tool 2.

It is observed that the results that both boys and girls had higher learning performance and lower cognitive load while using tool 2. This justifies the Neural Network model selection of tool 2 for many of the learners. The data is also analyzed on the aspect of the demographic area .The results on the basis of the demographics is shown in the Table 2

Table 2. Demographic comparision of performance for different visualization tools

	Urban	Sub Urban	Rural
Tool 2	85.5	87	85
Tool 1	72.5	71	74

It is noted that from the table 2 the performance of the student was better with Tool 2 compared to Tool 1 in all the demographic areas. This result also justifies the Neural Network model selection of Teaching Machine tool for most users in the study.

4.1 Analysis of Results – Individual Wise

The Neural Network system was able to predict the appropriate tool for reducing the cognitive load for about 24 students in a total of 36 students. This is evident from the higher performance level achieved by the learners using the tool chosen by the Neural network. It is also found that a strong relation exists between the minimum cognitive load and maximum performance for the majority of the students. The Neural Network model selected the correct tool for both concepts in the case of eleven students of the thirty six students. This is concluded considering the minimum cognitive load and

maximum performance for the tool chosen by the neural network. Another thirteen students expressed lower cognitive load and higher performance in the first concept using the visualization tool chosen by the Neural Network model. However, these thirteen students expressed lower cognitive load with the other tool for the second concept. On the contrary their learning performance was high with the tool chosen by the neural network model as in the case of first concept .The lower cognitive load could be due to the experiencing of using a new tool. In addition to this the student had gained the knowledge of the concept with the first tool which could also be the reason for the lower cognitive load. The learning performance was higher for both the concepts for the tool selected by the Neural Network .The Neural Network model is able to predict the appropriate tool for 24 out of 36 students which accounts to 67% accuracy. The accuracy of the neural network tool could be improved by additional training data. Thus the efficiency of the neural network will improve further.

The results were analyzed on the basis of the male learners, female learners, urban, semi urban and rural learners. Table 3 shows the percentage of the student population in male and female category having high performance and lower cognitive load while using the two ViLLE and Teaching Machine.

Table 3. Comparision between male and female in using the two tools

	Tool1		Tool2	
	Performance	Low CL	Performance	Low CL
Males	55	36	94	64
Females	47	47	53	53

It is observed from the above table 3 that both males and females experienced lower cognitive load and higher performance while using tool 2. The 61% of the male students experienced the lower cognitive load and had higher performance while using tool 2 and 50% of the female students experienced the lower cognitive load and higher performance while using tool 2. The following table 4 tabulates the comparison between the tool 1 and tool 2 among various demographic areas.

Table 4. Comparision between various demographics in using the two tools

	Tool 1		Tool 2	
	Higher Performance	Low CL	Higher Performance	Low CL
Urban	0		100	
Semi Urban	8	43	92	57
Rural	19	55	81	45

The Neural Network model has selected the Tool 2 as the appropriate tool in many cases and it matches with the higher performance and lower cognitive load of the majority of the learners while using the tool 2. The results validate the Neural Network

model in selecting the appropriate tool and also in establishing the relationship of inverse proportionality between the cognitive load and learning performance.

5 Conclusion and Future Work

The Neural Network model developed to choose the appropriate tool was successful in most cases. It is able to select the appropriate tool for almost 24 out 36 students in the study. The minimum cognitive load and maximum performance is experienced by the majority of the learners while using the tool selected by the neural network model. The model incorporates both the cognitive load and the LTM as the input with parameters such as gender, demographics and programming language as variables. Thus the proposed model could customize the instruction considering the various factors that impact the learning process by selecting the appropriate tool for learning. This model has been validated by checking the performance of the students in learning a concept using the tool chosen by Neural Network and also by allowing the learners to learn the same concept using the other tool which was not chosen by the neural network model. The result shows the overall performance was less in the tool 1 which was not mostly chosen for learning by the neural network model. The students could perform higher with the tool chosen by the neural network model. This model can be further improved with more training set data and continuous training. This framework could be incorporated to any to e learning environments to customize the instruction based on the profile and interest of the students. Thus it will help to optimize the nstruction to the learners and achieve intended learning outcomes.

References

1. Alkinani, M.: Software Visualization-Examination of its present status and its future applications (2009)
2. Amy, B.W., Tracy, C.G., Tai-Lung, C., Sherri, S.: Personality as a predictor for student success in programming principles I. Paper Presented at the 7th Annual Conference of Southern Association for Information System (2000)
3. Anderson, E.W., Potter, K.C., Matzen, L.E., Shepherd, J.F., Preston, G.A., Silva, C.T.: A user study of visualization Effectiveness using EEG and Cognitive Load. Paper Presented at the IEEE Symposium on Visualization (2011)
4. Antonenko, P., Pass, F., Grabner, R.: Using Electroencephalography to Measure CognitiveLoad. Educational Psychol. Rev. 22, 425–438 (2010)
5. Baddeley, A., Della, S.: Working Memory and Executive control. Philosophical-Transactions: Biological Sciences 351, 1397–1404 (1996)
6. Carter, J., Jenkins, T.: Gender differences in programming? Paper Presented at the 7th Annual Conference on Innovation and Technology in Computer Scienced Education (2002)
7. Ikehara, C.S., Crosby, M.E.: Assesing Cogntive Load with Physiologica Sensors. Paper Presented at the 38th Hawaii International Conference on System Sciences, Hawai, USA (2005)
8. du Boulay, B. (ed.): Some difficulties of learning programming. Lawerence Erlbaum, Hillsdale (1989)

9. Tuovinen, J.E.: Optimising Student Cognitive Load in Computer Education. ACM, Melbourn (2000)
10. Winslow, L.E.: Programming Pedagogy- A Psychological Overview. SIGCSE BULLETIN 28(3) (1996)
11. Essi, L., Kirsti, A.M., Hannu, M.J.: A study of the Difficulties of Novice Program mers. Paper Presented at the ITiCSE 2005. Monte de Caparica, Portugal (2005)
12. Garner, S.: A quantitative study of software tool that supports a Part complete solutionMethod on Learning Outcomes. Journal of Information Technology Education 8, 285–310 (2009)
13. Gomes, A., Mendes, A.J.: Learning to program-difficulties and solutions. Paper Presented at the International conference on Engineering Education - ICEE 2007, Coimbra, Portugal (2007)
14. Hart, S.G., Battiste, V., Chesney, M.A., Ward, M.M., McElroy, M.: Comparison of workload, performance and cardiovascular measures: Type A personalities vs Type B. Working paper. NASA Ames Research Center, Moffett Field (1986)
15. Hart, S.G.: NASA TLX Load Index (NASA TLX): 20 years later. Paper Presented at the Ergonomic Factors Society 50th Meeting, Santa Monica (2006)
16. Iain, M., Gleen, R.: Difficulties in Learning and Teaching Programming- Views of Students and Tutors. Education and Information Technologies 7(1), 55–66 (2002)
17. Ikehara, C.S., Crosby, M.E.: Assessing Cognitive Load with Physiological sensor. Paper Presented at the 38th Hawaii International Conference on System Sciences, Hawaii, USA (2005)
18. Kaastra, I., Boyd, M.: Designing a neural network for forecasting financial and economic time series. Neuro Computing 10, 215–236 (1996)
19. Kolb, D.A. (ed.): Learning Style Inventory: Technical Manual. McBer and company, Boston (1985)
20. Ala-Mutka, K.: Problems in Learning and Teaching Programming. Institute of Software systems, 13 (2007), http://www.cs.tut.fi/~edge/literature_study.pdf
21. Bruce-Lockhart, M.P., Norvell, T.S.: Lifting the Hood of the computer: Program Animation with the Teaching Machine. Paper Presented in Electrical and Computer Engineering, 2000 Canadian Conference (2000)
22. Paas, F., Juhani, E.T., Huib, T.: Cognitive Load Measurement as a means to Advance Cognitive Load Theory. Educational Psyhchologist 38(1), 63–71 (2003)
23. Pennington, N.: Comprehension strategies in programming. In: Olson, G.M., Sheppard, S., Soloway, E. (eds.) Emperical Studies in Programming, pp. 100–112 (1987)
24. Rajala, T., Laakso, M.J., Kaila, E., Salakoski, T.: ViLLE – Muti language Tool for Teaching Novice Programming TUCS Technical report. Helsinki: Tampere University, Finland (2007)
25. Rajala, T., Laakso, M.J., Kaila, E., Salakoski, T.: Effectiveness of program visualization – A case study with ViLLE Tool. Journal of Information Technology Education 7, 15–32 (2008)
26. Robins, A., Rountree, J., Rountree, N.: Learning and Teaching Programming: A Review and Discussion. Computer Science Education 13(2) (2003)
27. Schoneburg, E.: Stock Price Prediction Using Neural Networks: A Project Report. Neuro Computing 2, 17–27 (1990)
28. Shaffer, D., Doube, W., Tuovinen, J.: Applying Cognitive Load Theory to Computer Science Education. Paper Presented at the Joint Conference EASE&PPIG (2003)
29. Soloway, E. (ed.): What do novices know about programming?. Ablex, Norwood (1983)

30. Soloway, E.: Learning to Program = Learning to construct mechanism and explanations. Communications of ACM 29(9) (1986)
31. Soloway, E., Spohrer, J. (eds.): Studying the Novice programmers. L. Erlbaum Associates, New Jersey (1989)
32. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science 12, 257–285 (1988)
33. Wiedenbeck, S., Ramalingam, V., Sarasamma, S., Corritore, C.: A comparison of the comprehension of Object Oriented and procedural programs by novice programmers. Interaction with Computers 11(3), 255–282 (1999)
34. Wilkowski, W., Budzynski, T.: Application of Artificial Neural Networks for Real Estate Valuation. Paper Presented at the XXXIII FIG Congress, Munich, Germany (2006)
35. Winslow, L.E.: Programming pedagogy- A psychological overview. SIGCSE BULLETIN 28(3), 17–22 (1996)
36. Shi, Y., Choi, E.H.C.: Galvanic Skin Response (GSR) - as an index of Cognitive Load. Paper Presented at the CHI 2007, San Jose, CA, USA (2007)

Sequential Pattern Mining Application to Support Customer Care "X" Clinic

Alexander Setiawan, Adi Wibowo, and Samuel Kurniawan

Informatics Engineering Department of Petra Christian University
Siwalankerto 121-131, Surabaya – Indonesia
{alexander,adiw}@petra.ac.id, m26406143@john.petra.ac.id

Abstract. The X clinic that was one of the pioneers in the aesthetics clinic in Indonesia, had much experienced manpower. In a supplied manner this experience, the doctors and the nurse could to the X clinic serve consultations and could give the suggestion or the recommendation to the customer for the following maintenance. This matter gave comfort for the customer to take the decision. However, not all the customers had time that was enough to consult with the doctor. With used Free Span, one of the algorithms in the method sequential pattern mining was expected to be able to satisfy the requirement for the clinic customer of X. The use of sequential pattern mining in this recommendation system could help the doctor in increasing the recommendation, and helping the customer in taking the decision. This algorithm used the historic data the maintenance of the available customer. Results that were given in this program took the form of the pattern that in accordance with the available situation to the clinic of X. The result of the recommended selected based on existing customer categories, namely gender, priority customers in the clinic, and age range. Expected with the available category to be able to give the recommendation that agreed with the customer's available criterion.

Keywords: Sequential pattern mining, Free Span, Clinic.

1 Introduction

The X clinic was known as the pioneer in Aesthetic Clinic in Indonesia. Supplied with the experience for more than 16 years, the X clinic is currently known as one of the foremost aesthetics clinics in Indonesia with 11 branches that were spread in Jakarta, Surabaya, Denpasar, Kuta, Balikpapan, Malang, Batam and Makassar. With the "Beyond Result" concept, apart from giving results of the first-rate maintenance, the X clinic also gave comfort, the security in each one of his maintenance as well as the friendliness and privation for his customer.

Now the role of the internet is felt to be increasingly important and significant. Through this internet, the X Clinic wanted to give a special service for the customers in the X clinic by giving an application web the recommendation. This recommendation could be used by the nurse or the doctor to be able to give maintenance information what could be taken by the customer, and could be used by the customer as help in taking the decision.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 140–151, 2015.
DOI: 10.1007/978-3-662-46742-8_13

Therefore, with the existence of the data concerning historic the customer who carried out the maintenance in this clinic could be known by the maintenance what could be taken by the further customer. From this available data, also could be known how the place or the pattern of the maintenance that often was taken by the customer with used sequential pattern mining that was used to look for and find relations between item available to some of dataset. Specifically problems that was appointed in the production of this web was [5] :

- How to change from the long database (in this case used .dbf) became the new database (the database sql).
- How the algorithm that was used in the method sequential pattern mining.
- How processed the customer's data to the X clinic with used sequential pattern mining.
- How could the application give information concerning any maintenance that could be used in order with used sequential pattern mining.

2 Sequential Pattern Mining

Sequential pattern mining is a data mining technique used to identify patterns of ordered event within a database. First introduced in 1995 by Rakesh Agrawal of IBM's Almaden Research Center [1]. Was the method of look for and finding relations between item available to some dataset. The program sequential pattern mining aimed at finding information item-item that was connected with each other in the form of rule. Therefore sequential pattern exact to be applied in mining sequential pattern [2].

The available problem in the data mining was to find the pattern sequential. Inputan his data was a collection sequence that was mentioned data-sequences. Each data sequential was a list from transactions, where each transaction was a collection item. Generally a transaction was distinguished according to occurrence time of this transaction. Some sequential-pattern also consisted the list from a collection item [2], [3].

Anyway, all item in one a part of a pattern sequential must be gotten in a data transaction sequential to support this pattern. For the ease could diasumsikan that there is no data-sequence had more than one transaction with same transaction time, and used transaction time as the connoisseur of this transaction. So, in this case the quantity or the amount of some item was not counted on [3].

2.1 FreeSpan Algorithm

The FreeSpan Algorithm consisted of several stages, that is the first stage made sequence the database and searched f list or also could be mentioned as length-1 sequential pattern, the second stage made frequent item matrix, the third stage made the table pattern generation that consist of length-2 sequential pattern, item repeating pattern, and projected DB, and the last stage was the production sequential patterns [4].

The first stage from this algorithm was made sequence the database was based on the available data of the transaction data [4]. This data could be distinguished in accordance with the time and did not see the quantity from this transaction shown in Table 1.

Table 1. The Example Table of The Transaction

ID	Tanggal	Item
10	28/03/2010	b
10	28/03/2010	d
10	15/04/2010	c
10	23/05/2010	b
10	22/06/2010	a
10	22/06/2010	c
20	21/03/2010	b
20	21/03/2010	f
20	22/04/2010	c
20	22/04/2010	e
20	12/05/2010	b
20	11/06/2010	f
20	11/06/2010	g
30	08/04/2010	a
30	08/04/2010	a
...	...	...
50	13/06/2010	e

From the example of the data to the Table 1 could be made by one sequence the database. The production sequence the database made the time (in the example above, take the form of the date) as the reference in grouped sequence the database like the example to the Table 2. [2], [4].

FreeSpan Algorithm considering the pattern-projection method for mining sequential patterns. It is the original approach for mining sequential pattern recursive projecting the data sequence into smaller database [5].

Table 2. Tabel Sequence Database

ID	*SEQUENCE*	*ITEM_PATTERN*
10	(b,d),c,b,(a,c)	a,b,c,d
20	(b,f),(c,e),b,(f,g)	b,c,e,f,g
30	(a,h),(b,f),a,b,f	a,b,f,h
40	(b,e),(c,e),d	b,c,d,e
50	a,(b,d),b,c,b,(a,d, e)	a,b,c,d,e

From the Table 1 could be seen how the production of the Table 2 in the column sequence, contained the list of the available transaction with used the date as the arbiter. For the transaction that happened in the same date, his writing was distinguished by signaling (). For example will be explained by the production sequence by ID 10.

In the date 28/03/2010 bought the thing b and d, so as to be grouped to (b,d). In the date 15/04/2010 only bought c, so as only was written c. In the date 23/05/2010 the available transaction only bought the thing b, so as only was written b. In the date 22/06/2010 bought an and c, so as to be grouped to (a,c). From the data - this data had finally been grouped to one sequence, that is (b,d) ,c,b, (a,c).

After getting the table sequence, looked for the collection from the data that often emerged. This collection was also mentioned as length-1 sequential pattern. This collection will be put in order be based on support descending order, before looking for this collection determined previously minimal support. Minimal support for this example was 2. Length-1 sequential pattern in this example was as follows b: 5, c: 4, a: 3, d: 3, e: 3, f: 3. Length-1 sequential pattern this also could be mentioned frequent item list or more briefly f list.

The second stage that is made frequent item matrix F to count the frequency from the incident every time length-2 sequence that was formed be based on f list. To f list (i1,i2,... ,im), F was matrix the F triangle [j,k], where 1 j m. and 1 k j. F [j,j] (for 1 j m.) only had 1 counter, whereas the rest of them of F [j,k] (1 j m. and 1 k j) had 3 counters: (A,B,C), where A was the number of incidents where ik was after ij (for example in the table sequence was gotten < ij,ik >), B was the incident where ik was before ij (for example in the table sequence was gotten < ik,ij >), and C was the incident where ik happened together be the same as ij (for example in the table sequence was gotten < (ij,ik) >). Was based on this example, was gotten 6 available things in f list. This matter to make the triangle matrix with the measurement 6x6, with each value = 0, could be seen in the Table 3.

Table 3. Table Frequent Item Matrix

Item	Output length-2 sequential pattern	Ann. On Repeating Items	Ann. On Projected DB
f	<bf>:2;<fb>:2;<(bf)>:2;<ff>:2	{b^+f^+}, <ff^+>	Ø
e	<be>:3;<(ce)>:2	<b+e>	<(ce)>:{b}
d	<bd>:2;<db>:2;<(bd)>:2;<cd>:2;<dc>:2;<da>:2	{b^+d};<da^+>	<da>:{b,c};{cd}:{b}
a	<ba>:3;<ab>:2;<ca>:2;<aa>:2	<aa^+>:{a^+b^+};<ca^+>	<ca>:{b}
c	<bc>:4;<cb>:3	{b^+c}	Ø
b	<bb>:4	<bb^+>	Ø

The matrix above will be used to make the stage further, that is making the table pattern generation that consist of length-2 sequential pattern, item repeating pattern, and projected DB. The step to make the table pattern generation was as follows:

- Made length-2 sequential patterns: To every time counter from the matrix, if thought from counter was not smaller than minimal support, put forward frequent pattern this.
- made annotation ounce item repeating patterns for each column j: for the side diagonal, if F [j,j] min support, put forward < jj+ >. For the column i j, there was something that must be paid attention to: if F [i,i] min support then was contained i+, if F [j j] min support then was contained j+.

- made annotation ounce projected databases for each column j: To every time i < j, if F [i,j], F [k,j], and F [i,k] (k < i) can make pattern generating triple (with the note of each pair often emerged), k could be added in the column projected to i. After investigated all the columns in front i, the column projected could be determined.

The production of the table pattern generation for the example above could be seen in the Table 4.

Table 4. Table The Pattern Generation

b	4					
c	(4,3,0)	1				
a	(3,2,0)	(2,1,1)	2			
d	(2,2,2)	(2,2,0)	(1,2,1)	1		
e	(3,1,1)	(1,1,2)	(1,0,1)	(1,1,1)	1	
f	(2,2,2)	(1,1,0)	(1,1,0)	(0,0,0)	(1,1,0)	2
	b	**c**	**a**	**d**	**e**	**f**

3 Analysis and System Design

The X clinic that was the pioneer in Aesthetic Clinic in Indonesia has used the computer system in recording the transaction and his details base on website [6]. This system also recorded the data from this clinic customer. The customer who just the first time was present being asked for by the data himself, in part: the name, the address, gender, and the date of birth. After information was finished was recorded, the customer could choose the maintenance kind what will be taken and chose the doctor to be able to carry out consultations. For the customer who for a long time has been registered could make the promise or appointment with this clinic. This agreement contained the date and the hour for the maintenance or carried out consultations with the doctor. After carrying out the agreement, immediately will be carried out by the good action medically or non-medical. However the problem that was dealt with was several patients who came to experience the difficulty in choosing the maintenance kind, especially for the patient who just tried the maintenance. Where as the X clinic, still used the method of maintenance consultations in a traditional manner that is consulting directly with the doctor.

From the problem that often emerged, then could be concluded that the company needed a based system the computer that could help in decision making by using the story data of the maintenance of the available customer as the source of the beginning data. The system that was needed by the clinic was:

- The system was based on the data mining, that could produce information for the customer in carrying out decision making by using the story data of the maintenance of the available customer.

- The system that used the concept sequential pattern, where could show connection relations between the maintenance from various abstraction stages, not only from treatment that was taken simultaneously, but also in order.
- The computer system that could integrate all the factors of the customer's decision making to produce good information for the customer and for the clinic. The system used the concept of Frequent Pattern-Projected Sequential Pattern Mining (FreeSpan), where the factor or this decision making variable was: gender, range the age, the category of the customer, and the price of the maintenance.

The process mining that the first time was carried out was the process selection or the election. In this process was carried out by the type selection customer.

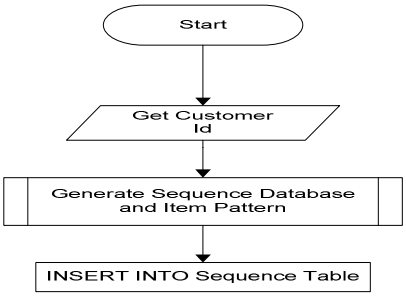


Fig. 1. Flowchart Selection Process

The selection process on figure 1. was gotten by 2 process sorts more, that is generate sequence the database and generate item pattern the Election was based on gender, range the age, the category of the customer, and the price of the maintenance. The process generate Item pattern this will be used to get length-1 sequential pattern or also could be mentioned as frequent item list (f list). took the data item pattern from the table sequence that will be used to search length-1 sequential pattern or often was mentioned frequent item list (f list). Frequent Item Matrix. Matrix that was produced from this process was very important his use in searching length-2 sequential patter. After that was carried out Generate Pattern Generation. This process used Frequent Item Matrix and the table sequence in his production. Results of this process of taking the form of 3 sorts, that is length-2 sequential pattern, annotation ounce repeating items, and annotation ounce projected databases. The flowchart generate pattern generation can be seen figure 2.

Along with this was the data Current in the system mining in the form of DFD that was the explanation from the figure 1 and figure 2. In DFD this was gotten by 6 processes that will be it was explained further that is the process selection, sorted frequentitems, frequent item matrix, generate pattern generation, generate rule, and translate rule can be seen in Figure 3.

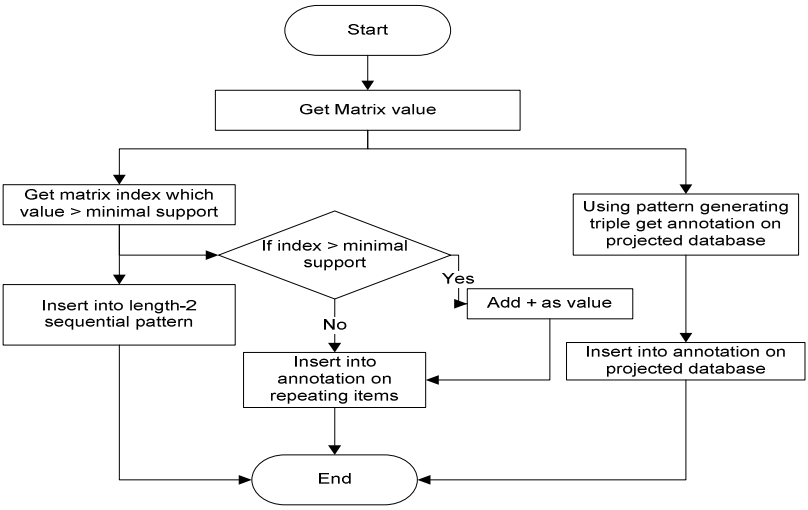


Fig. 2. Generate Pattern Generation

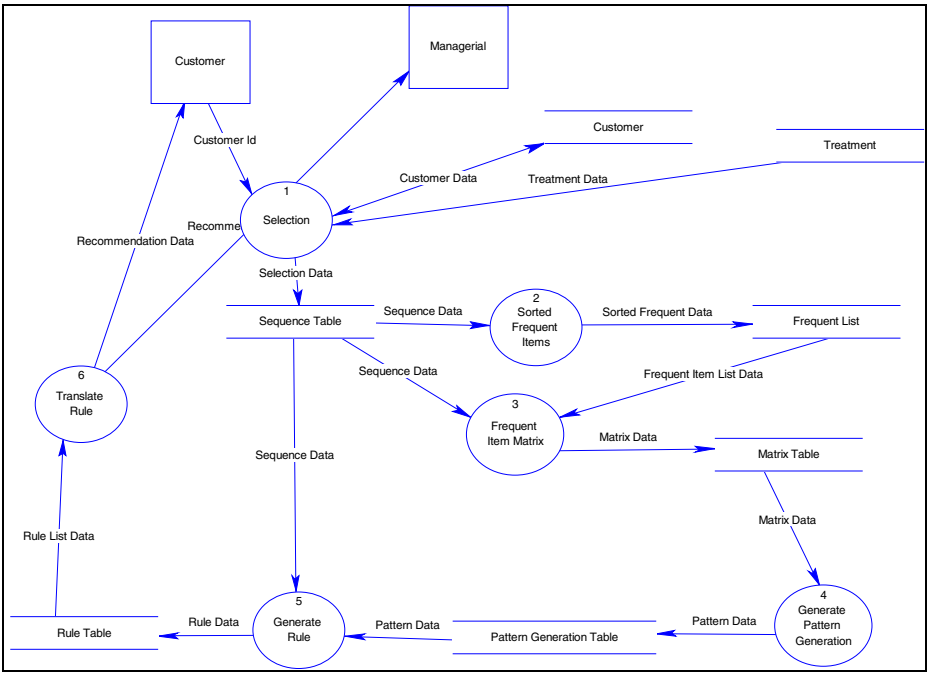


Fig. 3. Process Data Flow Diagram Mining Clinic “X

4 Implementation and Testing

This explained concerning the implementation to each controller together with available functions to each controller. Controller connected view and the model. Each controller had the function that differed to every time and the available process in this recommendation system.

4.1 Testing of Data Mining

The page of the management mining was used to get the data from the recommendation be based on historic the customer and the category who were given. The category that was used in getting the recommendation to be gender, the category customer, and range the age. The Home pages mining for representative of Aesthetic Clinic can be seen figure 4. To this page was gotten by main activity access from the clinic of "X".

Fig. 4. Home Pages Mining

To this page in figure 5. was gotten by the category that was used to get the recommendation data. The available category was gender, the priority in the customer, and range the age.

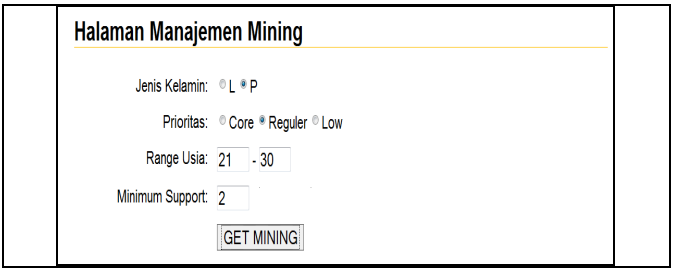


Fig. 5. Enter Mining Category

Gender that was used as the reference was L (The Man) and P (The Woman). The priority in the customer who was used was Core (the core customer), Regular (the pub-

lic's customer), and Low (the customer with the minimal priority). Range the age that was given here was used to restrict the historic data the customer who was used.

From the category that was given, the program will make the table mining to facilitate the process of the production sequential pattern by means of looking for the historic data the customer that in accordance with the category and put in order was based on the customer's data. The process was in the production of this table put in order be based on id from the customer. The column sequence was historic that was recorded from the customer with id this, and the column item pattern was the list from the maintenance that was recorded in historic this customer. The writing in the column sequence was adapted to time that was used by this customer. The mining sequence table can be seen figure 6.

Mining Sequence Table

ID	SEQUENCE
10029	LSCT01,LSPX01,LMKT01,LMMC01,LSCO13,RFFC03,LMKT01,(AMRJ02,BDBM05),CPSA01,LMKT01,OKBP01,LMMC01,L
10359	(LSEL02,LSMD10,MKCL04),(LLFT01,LNFH06),(AMWH01,FCWH01,MSFC25),ACVL013,ACVL003
11063	RFFC03,FCRJ06,RFFC03,LMKT01,(CPFC02,CPSA01),(LSMD11,RFFC03),(CPFC10,CPNC01),ECHF04,(ECHF02,INFL01)
11841	MKDP01,LMKT01,(FCAA02,LNNC01),LSEL01,LMKT01,BDBM08,(FCWH01,LNEY01,LNNC01)
11979	LSPF99,INWH02,INWH02,INWH02,INWH02,INWH02,INWH02,(INWH02,LSPF01),(AMRJ02,LLFT01,LNFH06),(INWH02,LSF (ACVL003,MSFC24),(INBT07,INBT09,INBT10),LSCT01,FCWH02,ILAC01,(ECHF04,INBT04,INBT12,LSCT01),LSCT01
1327	(LSPX01,RFFC03),(INBT01,INBT07,MKCL03,MKNC01,RFFC03),MKAC03,(INBT07,INBT12),(LSPX01,MKCL05,MKNC01,R (INBT07,MKWH03,RFFC03),(ILAC01,INBT09,MKWH03,RFFC03)
134	LSLP01,(CPFC07,MKNC01,RFFC03),(FCWH01,LNEY01,LNFH04,LNNC01,LNST01,MKCL04,MSLL99),(INBT04,INBT08),(C
15035	CPFC07,(FCWH01,LNST01),(AMRJ02,FCWH06,MKWH03)
15193	FCLL02,BDBM07
15198	LSCT01,LSCT01,LSCT01,LSCT01,LSCT01,LSPF99,LSCT01,MKEY01,(LNNC01,LSEL10,MSFC25),(BDBM16,MSBM08),IL
16621	(CPEY02,CPFC07,INBT06,INBT09,INWH02,LSCT01,MKCL04,MKNC01,MSBD02,RFFC03),(CPEY02,CPFC07,INWH02,LS (CPEY02,CPFC07,INWH02,LSCT01,LSEL01,MKCL01,MKWH03,RFFC03),(AMRJ02,DFLL99,ILAC01,INBT03,INBT07,LSM (ACGL003,ACGL039,ACGL054,ACGL056,ACGL063,ACGL064,ACGL065,ACGL066,ACGL68,ACGL69,ACVL016,ACVL019 (ACGL003,ACGL039,ACGL054,ACGL056,ACGL063,ACGL064,ACGL065,ACGL066,ACGL68,ACGL69,ACVL016,ACVL019
170	LSEL12,(AMRJ02,FCWH01,LLFT01,LNEY01,LNFH06,LNNC01,LNPR01,LNPR02,LNST01,MKNC01),(AMRJ02,FCWH06,LN (AMRJ02,FCWH06,LNEY01,LNFH05,LNNC01,LNST01,MKWH03,MSLL99),(FCWH01,LNEY01)

Fig. 6. Form Mining Sequence Table

After getting the mining sequence table, the process will be followed by the Matrix production after got frequent item list or length-1 sequential pattern from the column item pattern, the form matrix table can be seen figure 7.

Mining Sequence Table > Matrix

5															
4,5,2	10														
5,1,1	5,4,0	5													
4,6,1	5,7,0	4,6,3	9												
4,3,1	3,3,0	5,4,3	5,2,2	8											
7,5,8	5,3,3	2,4,0	4,3,0	5,5,1	11										
8,6,7	5,3,1	4,6,0	4,4,0	3,4,1	9,9,11	8									
7,5,8	4,3,1	3,4,1	2,3,1	4,4,0	6,8,7	8,7,9	10								
2,5,2	4,5,0	2,8,1	5,3,1	2,3,3	1,4,0	2,2,1	1,3,1	3							
3,4,7	3,1,1	1,2,0	1,3,0	3,4,0	3,6,8	2,6,8	5,8,10	2,0,0	4						
2,4,1	2,1,0	1,1,0	1,2,0	2,2,0	4,4,6	3,3,7	3,2,4	1,0,0	3,0,4	2					
1,4,0	1,3,1	3,4,2	4,3,3	1,5,2	2,3,1	3,4,1	2,3,1	4,4,3	3,3,1	1,2,0	6				
2,3,0	5,4,0	1,5,0	7,2,0	0,3,0	0,2,1	1,1,0	1,1,0	6,3,0	0,1,1	0,0,0	4,2,2	3			
2,5,0	3,3,0	0,1,0	2,3,2	3,2,1	2,3,0	2,3,0	2,0,0	3,2,1	2,0,0	2,4,0	3,3,2	1,3,0	3		
2,3,0	2,4,0	1,3,7	3,3,3	0,5,2	0,1,0	1,1,0	1,0,0	3,2,1	1,0,0	0,0,0	2,2,2	3,2,0	1,2,1	1	
3,4,5	2,2,0	0,2,0	1,2,0	2,2,0	2,6,7	3,4,5	4,5,5	2,1,1	3,2,5	1,2,4	2,1,1	2,1,1	3,1,0	0,0,0	
1,1,0	2,2,0	3,5,1	4,3,2	1,1,0	2,2,0	2,1,0	0,1,0	3,2,2	0,1,0	1,1,0	3,2,2	3,2,0	2,0,0	3,2,1	
6,4,3	3,1,0	5,3,1	5,2,2	2,2,1	4,2,3	5,4,3	4,3,2	3,3,2	3,0,1	3,1,1	2,1,2	3,2,0	3,2,1	3,1,1	

Fig. 7. Form Matrix Table

From matrix that was formed, the process mining will be continued by searching length-2 sequential pattern that will produce repeating item pattern and projected

Mining Sequence Table > Projected DataBases

ANNOTATION	ID	SEQUENCE
{LMKT01 INBT07} : {AMRJ02}	10029	LSCT01,LSPX01,LMKT01,LMMC01,LSCO13,RFFC03,LMKT01,(AMRJ02,BDBM05
{LMKT01 INBT07} : {AMRJ02}	4686	(LSLP05,LSMD11),ILAC01,LSLP05,LMKT01,LMKT01,LMKT01,LMKT01,LMKT01,(A (ACVL004,ACVL011,LNEY01,LNNC01,LSCO13,MKWH04),LMKT01
{LMKT01 INBT07} : {AMRJ02}	7413	(LLBA05,LSMD10),XXXX99,(INFL02,LSMD10),LMKT01,LSMD10,LMKT01,(LSMD1 (AMRJ02,AMWH01,LNEY01,LNNC01,LNST01,MKNC01,MKWH03),INFL01,(AMWH (ACVL004,AMWH01,MKWH03,MSFC24)
{LMKT01 INBT07} : {AMRJ02}	8873	(INBT02,INBT07,INBT09,INFL01,INFL04,INWH02,RFFC03),(CPSA01,INWH02,LSEL (INBT02,INBT04,INFL04)
{LMKT01 RFFC03} : {AMRJ02}	10029	LSCT01,LSPX01,LMKT01,LMMC01,LSCO13,RFFC03,LMKT01,(AMRJ02,BDBM05
{LMKT01 RFFC03} : {AMRJ02}	18211	LMKT01,CPFC06,ILAC01,LSEL12,(FCAC07,LMKT01,LNFH04,LNNC01,MKDP01),(((AMRJ02,LNEY01,LNNC01,MSFC25),RFFC03,(AMST01,LNEY01,LNFH05,LNNC01
{LMKT01 RFFC03} : {AMRJ02}	275	LMKT01,MKCL02,LSPX01,CPFC05,(LNFH06,MKCL03,RFFC03),(MKCL05,RFFC03 (ACVL003,ACVL004,AMRJ02,FCRJ02,LNEY01,LNFH06,LNFH08,LNFH11),DFLL99
{LMKT01 RFFC03} : {AMRJ02}	8873	(INBT02,INBT07,INBT09,INFL01,INFL04,INWH02,RFFC03),(CPSA01,INWH02,LSEL (INBT02,INBT04,INFL04)
{INBT07 RFFC03} : {LMKT01,AMRJ02}	10029	LSCT01,LSPX01,LMKT01,LMMC01,LSCO13,RFFC03,LMKT01,(AMRJ02,BDBM05
{INBT07 RFFC03} : {LMKT01,AMRJ02}	16621	(CPEY02,CPFC07,INBT06,INBT09,INWH02,LSCT01,MKCL04,MKNC01,MSBD02,R (CPEY02,CPFC07,INWH02,LSCT01,LSEL01,MKCL01,MKWH03,RFFC03),(AMRJ0 (ACGL003,ACGL039,ACGL054,ACGL056,ACGL063,ACGL064,ACGL065,ACGL06 (ACGL003,ACGL039,ACGL054,ACGL056,ACGL063,ACGL064,ACGL065,ACGL06

Fig. 8. Projected Database Mining

the database. In this program, the process will be continued to Projected Database in accordance with that was put forward to the figure 8.

From the data that was received in projected the database, could be looked for the pattern with combined projected the database. And results of this combination were used to get sequential pattern and support him can be seen figure 9.

Mining Sequence Table > Sequential Pattern

ANNOTATION	SEQUENTIAL PATTERN	SUPPORT
{LNEY01 FCWH01} : {LNNC01,ILAC01,RFFC03}	(FCWH01,LNEY01,LNNC01)	6
{LMKT01 LNEY01} : {AMRJ02}	LMKT01,(AMRJ02,LNEY01),LNEY01	4
{LMKT01 LNEY01} : {AMRJ02}	LMKT01,AMRJ02,LNEY01	4
{LMKT01 LNEY01} : {AMRJ02}	LMKT01,(AMRJ02,LNEY01)	4
{LNNC01 LNEY01} : {ILAC01,RFFC03,INBT07,LMKT01,AMR	(AMRJ02,LNEY01),(LNEY01,LNNC01)	4
{LNNC01 LNEY01} : {ILAC01,RFFC03,INBT07,LMKT01,AMR	(AMRJ02,LNEY01),LNNC01	4
{LNNC01 LNEY01} : {ILAC01,RFFC03,INBT07,LMKT01,AMR	(AMRJ02,LNEY01,LNNC01),LNEY01	4
{LNNC01 LNEY01} : {ILAC01,RFFC03,INBT07,LMKT01,AMR	AMRJ02,(LNEY01,LNNC01)	4
{LNNC01 LNEY01} : {ILAC01,RFFC03,INBT07,LMKT01,AMR	(AMRJ02,LNEY01,LNNC01)	4
{LNNC01 LNEY01} : {ILAC01,RFFC03,INBT07,LMKT01,AMR	(AMRJ02,LNNC01),LNEY01	4
{LNNC01 LNEY01} : {ILAC01,RFFC03,INBT07,LMKT01,AMR	AMRJ02,LNNC01,LNEY01	4
{LNFH06 AMWH01} : {LNEY01,LNNC01}	(LNEY01,LNFH06),AMWH01	4

Fig. 9. Form Sequential Pattern

Results of the end that was obtained were the pattern from sequential pattern that was interpreted. Results of this end were the reference in giving the recommendation to be based on the category that was determined in the beginning of the process of the management mining this.

The figure 10. put forward the example produced by the end that was received after carrying out the process mining towards the data customer that was matched with the category that was chosen at the beginning of the process.

Mining Sequence Table > Rule

SEQUENTIAL PATTERN	RULE	SUPPORT
(FCWH01,LNEY01,LNNC01)	(MRC MESODERMIC WHITENING FACIAL->MRC COLLAGEN EYE TREAT+LIFTING->MRC OXYGEN NECK TREATMENT)	6
LMKT01,(AMRJ02,LNEY01),LNEY01	MRC KONTROL=>(MRC AMPUL BIO COLLAGEN->MRC COLLAGEN EYE TREAT+LIFTING)=>MRC COLLAGEN EYE TREAT+LIFTING	4
LMKT01,AMRJ02,LNEY01	MRC KONTROL=>MRC AMPUL BIO COLLAGEN=>MRC COLLAGEN EYE TREAT+LIFTING	4
LMKT01,(AMRJ02,LNEY01)	MRC KONTROL=>(MRC AMPUL BIO COLLAGEN->MRC COLLAGEN EYE TREAT+LIFTING)	4
(AMRJ02,LNEY01),(LNEY01,LNNC01)	(MRC AMPUL BIO COLLAGEN->MRC COLLAGEN EYE TREAT+LIFTING)=>(MRC COLLAGEN EYE TREAT+LIFTING->MRC OXYGEN NECK TREATMENT)	4
(AMRJ02,LNEY01),LNNC01	(MRC AMPUL BIO COLLAGEN->MRC COLLAGEN EYE TREAT+LIFTING)=>MRC OXYGEN NECK TREATMENT	4

Fig. 10. Form Final Result Mining

4.2 Testing of Recommended User

To get the recommendation, member will be asked for to put id member before. Id member this was used to get the data himself member and was compared with the available category to the recommendation table. The data himself who could be taken was gender, the priority, and the date of birth. This page in figure 11. put forward results of the recommendation that was taken from the database with the data himself the customer with id 11979. The data himself who was taken repeatedly was the category that was needed, and the last maintenance that was carried out. In these results was seen that the last maintenance that was carried out by the customer had "MRC CONTOURING BODY TREATMENT MCBT", because that was based on the available data the recommendation that was first that is taking "MRC RF LIFT" and could be followed by "MRC INJEKSI ACNE 1".

Member Page

	PERAWATAN TERAKHIR	PERAWATAN REKOMENDASI	PERAWATAN REKOMENDASI LANJUTAN
Alternatif 1	MRC INJEKSI ACNE 1 =>	MRC AMPUL BIO ACNE dan MRC ACNE BIO FACIAL =>	
Alternatif 2	MRC INJEKSI ACNE 1 =>	MRC KONTROL =>	MRC BETA PEEL =>
Alternatif 3	MRC INJEKSI ACNE 1 =>	MRC AMPUL BIO ACNE dan MRC ACNE BIO FACIAL =>	MRC INJEKSI ACNE 1 =>
Alternatif 4	MRC INJEKSI ACNE 1 =>	MRC AMPUL BIO ACNE dan MRC ACNE BIO FACIAL dan MRC INJEKSI ACNE 1 =>	
Alternatif 5	MRC INJEKSI ACNE 1 =>	MRC KONTROL =>	MRC BETA PEEL =>

'=>' perawatan berikutnya

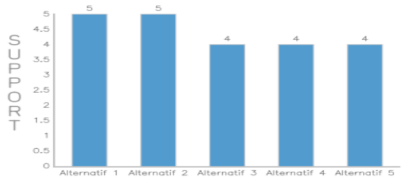


Fig. 11. Form Recommendation Member

5 Conclusion and Advise

Was based on available discussions in chapters beforehand then could be taken by several conclusions as follows:

- From the test with the data that actually, could be produced the pattern of the pattern that could be recommended to the customer. The pattern of the pattern that was gotten in these results was made be based on FreeSpan Algorithm in the method sequential pattern mining. The source of the data that was used in this algorithm, was processed in accordance with the category that is gender, the priority in the customer, and range the age.
- Through website that was made, the Administrator could get the pattern of the pattern of the maintenance that was taken by the customer in accordance with range the age in the category that was given.
- From the process of the testing on the whole, this program was it was felt more inappropriate in the search for the maintenance recommendation to the customer. Because the taking of the maintenance that was carried out by the customer often was carried out on the basis of the recommendation from the doctor, so as the pattern of the recommendation that was given not more maximal.
- Must be tried to develop this recommendation system with more made his execution time efficient and maximized results that were received.
- Facilities in this application were still limited so as big the opportunity of increasing and increasing facilities to this site so that the achievement of this site could be more optimal.

References

1. Agrawal, R., Ramakrishnan, S.: Mining sequential patterns. In: Proc. Int. Conf. Data, pp. 3–14 (1995)
2. Han, J., Kamber, M.: Datamining: Concept & Technique, 3rd edn. Morgan Kaufmann Publisher is an imprint of Elsevier (2012)
3. Wright, A.P., Wright, A.T., McCoy, A.B., Sittig, D.F.: The use of sequential pattern mining to predict next prescribed medications. Journal of Biomedical Informatics (2014)
4. Han, J., Pei, J.: FreeSpan: Frequent pattern-projected sequential pattern mining (2010), http://www.cs.sfu.ca/~jpei/publications/freespan.pdf (retrieved April 7, 2010)
5. Desai, N., Ganatra, A.: Draw Attention to Potential Customer with the Help of Subjective Measures in Sequential Pattern Mining (SPM) Approach. In: Proceeding of International Conference on Recent Trends in Information, Telecommunication and Computing, ITC (2014)
6. Myer, T.: Professional CodeIgniter. Wiley Publishing (2008)

Similarity-Based Models

The Comparation of Distance-Based Similarity Measure to Detection of Plagiarism in Indonesian Text

Tari Mardiana, Teguh Bharata Adji, and Indriana Hidayah

Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada, Yogyakarta, Indonesia
{tari.mardiana,adji.tba}@gmail.com, indriana.h@ugm.ac.id

Abstract. The accesible loose information through the Internet leads to plagiarism activities use the copy-paste-modify practice is growing rapidly. There have been so many methods, algorithm, and even softwares that developed till this day to avoid and detect the plagiarism which can be used broadly unlimited on a certain subject. Research about detection of plagiarism in Indonesian Language develop day by day, although not significant as English Language. This paper proposes several models of distance-based similarity measure which could be used to assess the similarity in Indonesian text, such as Dice's similarity coefficient, Cosine similarity, and Jaccard coefficient. It implemented together with Rabin-Karp algorithm that common used to detect plagiarism in Indonesian Language. The analysis technique of plagiarism is fingerprint analysis to create fingerprint document according to n-gram value that has been determined, then the similarity value will be counted according to the same number of fingerprint between texts. Small data text about Information System tested in this case and it divided into four kinds of text document with some modified. First document is original text, second is 50% of original text adding with 50% of another text, third 50% original text modified using sinonym and paraphase, fourth some position of text in original text changed. From the experimental result, cosine similarity show better performance in generating value accuracy compared to the dice coefficient and Jaccard coefficient. This model is expected to be used as an alternative type of statistical algorithms that implement the n-grams in the process especially to detect plagiarism in Indonesian text.

Keywords: Fingerprint, Indonesian, Plagiarism, Similarity, Text.

1 Introduction

Information retrival through search engine grows rapidly as well as the growth of Internet usage in daily life. Information application in a free way sometimes create a higher fraud cases, such as plagiarism. A duplication activity or plagiarism is an act of duplicating some or whole parts of people's work especially written text without citing any notes of source (copyright) completely. In academic world, plagiarism is a common case for students do any assignments, reports, paper, even thesis that should be done by their own words instead of copy-paste modify. There are still a lot

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 155–164, 2015.
DOI: 10.1007/978-3-662-46742-8_14

of students who do not understand the difference between citation and plagiarism so that they do not see the edge between those things. Plagiarism practice is an unvoidable thing both in academic and daily lives, but it could be avoided by early detection based on the similarity level in the text, we could see in the grammar structure and the writing style.

Plagiarism detection through the right similarity measure is not only leading to excellent information but also should be great for lessen budget and time process. In the other hand, similarity measure technique could be divided by three kind of it [1], which are:

- Distance-based similarity measure, is a technique built for measuring the similarity level between two objects from geometric distance side of variables within those two objects. For example: Minkowski Distance, Manhattan/City block distance, Euclidean distance, Chebyshev distance, Jaccard distance, Dice's Coefficient, Cosine similarity, Hamming distance, and Soundex distance.
- Feature-based similarity measure, is a technique that built for measuring the similarity level by representing object into the feature that would like to compare, it is used to classify or pattern matching pictures and texts.
- Probabilistic-based similarity measure, is a technique built to measure the similarity level by representing two set of objects that are being compared in a probability form. For example: Kullback Leibler Distance and Posterior Probability.

This paper will discuss some measurement statistic of similarity level that could be implemented in analysis technique of plagiarism. The technique that will be discussed is fingerprint analysis uses Rabin-Karp algorithm. The objective of this experimental that it could be used the reference of the right statistic usage by matching along with the algorithm and process within it, including the use of n-gram in hash function. This research wished to give the right parameter value to be applied in the next research but not discussing about the complexity of time processing used.

2 Plagiarism Detection

In this section, we will explain about plagiarism analysis technique and algorithm used in research and some related work in similarity measure.

2.1 Rabin-Karp Algorithm

There are several techniques that could be used to analyze plagiarism, they are substring matching, keyword similarity, and fingerprint analysis [2]. Fingerprint Analysis is a very often used analysis approach for plagiarism with dividing documents into sequence principle called chunks. The pattern in documents will be changed with function hash as values represent string within them. Each string in the text will be changed using hash function for creating fingerprint document according to n-gram value that has been determined, then the similarity value will be counted based on the same fingerprint number between texts. The approach is used in this technique such as Rabin-Karp algorithm and Winnowing [3].

Rabin-Karp algorithm is a multiple pattern search algorithms that are very efficient for searching with the pattern of many [4]. This algorithm is often used for detection of pagiarism, especially in the Indonesian language text documents[5][6][7]. Disadvantages of this algorithm is a complex computational requirements and processing time that tends to be longer due to long influenced the content of the text document to be processed, starting from the stage of pre-processing and stemming that goes in it. However, the level of accuracy that is produced by this algorithm is high enough to detect plagiarism documents.

Rabin Karp algorithm using a hash function that simplifies the method to avoid the complexity of processing time in most cases. The principle of this algorithm is to streamline the examination only if the text is in the process are similar as in the pattern rather than conduct an examination of every position of the text when the pattern matching [7]. Hash function is a way to create a fingerprint of a variety of data inputs. Hash function will replace or transposing data to create the fingerprint, which is called the hash values. Hash value is described as a short string consisting of letters and numbers that look random (binary data written in hexadecimal notation). Forming hash value from string i and base d written in general as follow:

$$H(i) = i[0]d^{n-1} + i[1]d^{n-2} + ...+ i[n-1] \tag{1}$$

For example pattern "LUAS" with a base = of 10 have hash value:

$Hash_{(luas)}$ = [(107*10^3) + (117*10^2) + (97*10^1) + (115*10^0)]

= (107000+11700+970+115)

= 119785

In Rabin-Karp algorithm, hashing used for shifting substring search. If two strings are equal between the search string (m) with substring in the text (n), their hash values must equal too. Efficient calculation of the hash value at the time of the shift will affect the performance of this algorithm[8]. Pseudocode of Rabin-Karp algorithm can be written as follows:

```
function Rabin Karp(string s[1..n],
string sub[1..m])
hsub := hash(sub[1..m])
  for i from 1 to n-m+1
    if hs = hsub
      if s[i..i+m-1] = sub
      return i
      hs := hash(s[i+1..i+m])
    return not found
```

Fig. 1. Pseucode Rabin-Karp Algorithm

2.2 Distance-Based Similarity Measure

Similarity is a level comparison of similarity percentage between documents that are tested. The similarity number of percentage would be influenced by the similarity level of the tested document. The bigger similarity percentage the more similar the text would be[5]. Distance based-Similarity Measure is a technique built to measure the similarity level two objects by geometric distance side of variables include within the two objects [6]. Many theories assume that nearby objects would be more similar than far distant objects. Similarity measure (s) can be differentiated by distant measure (d) using quotation:

$$s = 1 - d \tag{2}$$

The most common unsimilarity counting process is leading to the distant measurement in metric space. in a information retrival system, the example of this approach includes the models or methods such as Minkowski Distance, Manhattan/City block distance, Euclidean distance, Chebyshev distance, Jaccard distance, Dice's Coefficient, Cosine similarity, Hamming distance, Levenshtein Distance, and Soundex distance [6]. The application of of similarity measurement model is not only used to detect the similar words but also able to use in chemical informatics [9].

2.2.1 Dice's Coefficient

A most common used measurement model to calculate similarity value with n-gram approach. Dice coefficient defines similarity level as twice as many common entity (n_t) divided by the two entities as total that are tested (n_x+n_y)[1]. The result value has range of 0 till 1 with coefficient 1 shows identical vector, and 0 equal to orthogonal vector [11].

$$\boldsymbol{s} = \frac{\boldsymbol{2n_t}}{\boldsymbol{n_x + n_y}} \tag{3}$$

Dice's similarity has the same form with F1-score, with monotonic nature within Jaccard similarity[9]. The other name of this model is Czekanowski or Sørenson coefficient. Measurements on the similarity value of research [3][6][12] using Dice's coefficient, which simply uses the same hash calculation amount divided by the total number of hash in the document.

2.2.2 Cosine Similarity

It is also known as Ochiai coefficient. It is a similarity measurement model based on vector that is commonly popular in the text mining and information retrieval. This approach compares string that has been changed into space vector so that cosinus in the Euclidean principle could be used to measure the similarity level [1].

$$\boldsymbol{similarity} = \boldsymbol{\cos\theta} = \frac{\vec{\boldsymbol{A}}.\vec{\boldsymbol{B}}}{\|\vec{\boldsymbol{A}}\|\|\vec{\boldsymbol{B}}\|} \tag{4}$$

2.2.3 Jaccard Coefficient

Jacard Coefficient or well known as Tanimoto coefficient as well, is a very common model to use in chemical informatics, whether to compare similarity, unsimilarity, and the distance between data set. The concept of this model is uses the ratio of the intersecting set to the union set as the measure of similarity [13]**.**

$$J_\delta(A,B) = \frac{|A \cap B|}{A + B - |A \cap B|} \tag{5}$$

where A and B represents the number attribute in each object, and $|A \cap B|$ is intersection set.

The benefit of using this model is it could show the close relation between two set data efficiaently without redudancy among the used data. This model will show the high precision result on the system and database in a smaller scale than implementing it on search page type with different service[14].

2.3 Similarity Using Fingerprints

The most common similarity measurement is calculated based on the document fingerprint. In Example given two text A and B below:

Text A → Suka itu belum tentu cinta
Text B → Saya suka kamu

Forming hash value from texts using hash function, and get C as matching hash value between A and B.

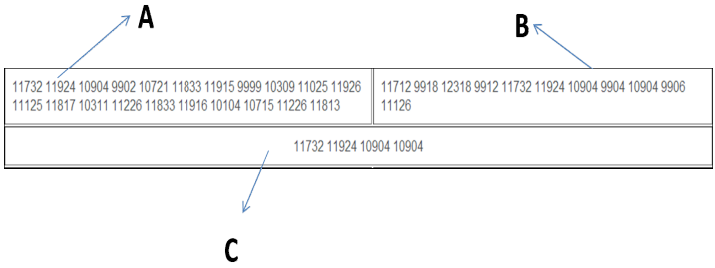


Fig. 2. Forming hash value from text

Simple similarity between Text A and B calculated using the model of distance-based similarity measure based on equations (3), (4), dan (5) as follows.

- *Dice coefficient*

$$Similarity = \frac{2C}{A + B} = 8 / (21 + 11) = 0.25$$

- *Cosine coefficient*

$$similarity = \frac{C}{\sqrt{(A \times B)}} = 4 / \sqrt{(21 \times 11)} = 0.263$$

- *Jaccard coefficient*

$$similarity = \frac{C}{A + B - C} = 4 / (21 + 11 - 4) = 0.142$$

Advantage in using hash values to search similarity between two text is string matching can be done regardless of value position.

3 Experimental Method

To determine the best models to detect plagiarism based on word similarity measurement, there are some process flow and test scenarios. Three types of models of distance-based similarity measure will be implemented as the statistics in the Rabin-Karp algorithm. Data text about Information System tested in this case and it divided into four kinds of text document with some modified. First document is original text, second is 50% of original text adding with 50% of another text, third 50% original text modified using sinonym and fourth some position of text in original text changed. Example of text used will be presented as follows:

a. Original Text

> Sistem Informasi (SI) adalah kombinasi dari teknologi informasi dan aktivitas orang yang menggunakan teknologi itu untuk mendukung operasi dan manajemen.Dalam arti yang sangat luas, istilah sistem informasi yang sering digunakan merujuk kepada interaksi antara orang, proses algoritmik, data, dan teknologi.
>
> Dalam pengertian ini, istilah ini digunakan untuk merujuk tidak hanya pada penggunaan organisasi teknologi informasi dan komunikasi (TIK), tetapi juga untuk cara di mana orang berinteraksi dengan teknologi ini dalam mendukung proses bisnis.
>
> Ada yang membuat perbedaan yang jelas antara sistem informasi, dan komputer sistem TIK, dan proses bisnis. Sistem informasi yang berbeda dari teknologi informasi dalam sistem informasi biasanya terlihat seperti memiliki komponen TIK. Hal ini terutama berkaitan dengan tujuan pemanfaatan teknologi informasi. Sistem informasi juga berbeda dari proses bisnis. Sistem informasi membantu untuk mengontrol kinerja proses bisnis.
>
> Sistem informasi adalah suatu sistem di dalam suatu organisasi yang mempertemukan kebutuhan pengolahan transaksi harian, mendukung operasi, bersifat manajerial dan kegiatan strategi dari suatu organisasi dan menyediakan pihak luar tertentu dengan laporan-laporan yang diperlukan.

Fig. 3. Original Text

b. Modification 1: 50% of original text adding with 50% of another text. Some sentences adding from another text but in the same field.

Sistem Informasi (SI) adalah kombinasi dari teknologi informasi dan aktivitas orang yang menggunakan teknologi itu untuk mendukung operasi dan manajemen.Dalam arti yang sangat luas, istilah sistem informasi yang sering digunakan merujuk kepada interaksi antara orang, proses algoritmik, data, dan teknologi.

Dalam pengertian ini, istilah ini digunakan untuk merujuk tidak hanya pada penggunaan organisasi teknologi informasi dan komunikasi (TIK), tetapi juga untuk cara di mana orang berinteraksi dengan teknologi ini dalam mendukung proses bisnis.

Alter berpendapat untuk sistem informasi sebagai tipe khusus dari sistem kerja. Sistem kerja adalah suatu sistem di mana manusia dan/atau mesin melakukan pekerjaan dengan menggunakan sumber daya untuk memproduksi produk tertentu dan/atau jasa bagi pelanggan. Sistem informasi adalah suatu sistem kerja yang kegiatannya ditujukan untuk pengolahan (menangkap, transmisi, menyimpan, mengambil, memanipulasi dan menampilkan) informasi.

Dengan demikian, sistem informasi antar-berhubungan dengan sistem data di satu sisi dan sistem aktivitas di sisi lain. Sistem informasi adalah suatu bentuk komunikasi sistem di mana data yang mewakili dan diproses sebagai bentuk dari memori sosial. Sistem informasi juga dapat dianggap sebagai bahasa semi formal yang mendukung manusia dalam pengambilan keputusan dan tindakan.

Fig. 4. Modified with adding sentences from another text

c. Modification 2: 50% original text using sinonym and paraphase. Word in bold is modification using synonim such as kombinasi replaced by perpaduan. Sentences in bold italic is modification using own words.

Sistem Informasi (SI) merupakan **perpaduan** dari teknologi informasi dan aktivitas **manusia** yang menggunakan teknologi tersebut untuk mendukung operasi dan manajemen. ***Istilah sistem informasi dalam pengertian yang lebih luas*** sering digunakan untuk merujuk kepada interaksi antara orang, proses algoritmik, data, dan teknologi.

Pengertian lain sistem informasi adalah suatu sistem dalam sebuah organisasi yang mempertemukan kebutuhan pengolahan transaksi harian, mendukung operasi, bersifat manajerial dan kegiatan strategi dari suatu organisasi serta menyediakan pihak luar informasi tertentu melalui laporan-laporan yang diperlukan.

Fig. 5. Modified with synonim and paraphase

d. Modification 3: Change some sentences position of 50% original text without change the meaning of text.

Sistem informasi digunakan untuk merujuk tidak hanya pada penggunaan organisasi teknologi informasi dan komunikasi (TIK), tetapi juga untuk cara di mana orang berinteraksi dengan teknologi ini dalam mendukung proses bisnis. Dalam pengertian ini, istilah ini adalah suatu sistem di dalam suatu organisasi yang mempertemukan kebutuhan pengolahan transaksi harian, mendukung operasi, bersifat manajerial dan kegiatan strategi dari suatu organisasi dan menyediakan pihak luar tertentu dengan laporan-laporan yang diperlukan.

Dalam arti yang sangat luas, istilah sistem informasi yang sering digunakan merujuk kepada interaksi antara orang, proses algoritmik, data, dan teknologi. Sistem informasi membantu untuk mengontrol kinerja proses bisnis.

Fig. 6. Modified with change some sentences position

4 Experimental Results and Analysis

Testing is done by comparing the text files with each other by changing n-gram value used in the algorithm. The recomendation selected base is 25 or higher which produces a constant value of the similarity. Original text will be compared with another modified texts and measure the similarity using Dice, Jaccard, and cosine coefficient.

Table 1. Comparation similarity measure result in indonesian texts (n-gram = 2)

Text	Similarity Measure	Text			
		Original	Modif 1	Modif 2	Modif 3
Original	Dice's	100	87.36	85.02	94.74
	Jaccard	100	77.56	73.93	90.00
	Cosine	100	87.37	85.40	94.84
Modif 1	Dice's	87.36	100	77.48	85.06
	Jaccard	77.56	100	63.24	74.00
	Cosine	87.37	100	77.96	85.23
Modif 2	Dice's	85.02	77.48	100	86.17
	Jaccard	73.93	63.24	100	75.71
	Cosine	85.40	77.96	100	86.27
Modif 3	Dice's	94.74	85.06	86.17	100
	Jaccard	90.00	74.00	75.71	100
	Cosine	94.84	85.23	86.27	100

Table 1 is a test scenario using n-gram = 2 to forming hash value according gram set. For the similarity measure used can be seen that the model with the cosine similarity distance measurements showed better performance in generating value accuracy, give small difference accuracy from dice's coefficient. This model is expected to be used as an alternative type of statistical algorithms that implement the

n-grams in the process. The result between Original text compared with modified text by change sentences position got the best result above 90% high accuracy. Basically, with the fingerprint similarity search based on the corresponding hash value, so it does not depend on the position of the words in the search for similarity called position independence [12].

Table 2. Comparation similarity measure result in indonesian texts (n-gram = 3)

Text	Similarity Measure	Text			
		Original	Modif 1	Modif 2	Modif 3
Original	Dice's	100	68.31	69.91	86.82
	Jaccard	100	51.87	53.76	76.70
	Cosine	100	68.35	71.14	87.47
Modif 1	Dice's	68.31	100	56.49	67.22
	Jaccard	51.87	100	39.37	50.63
	Cosine	68.35	100	57.89	68.05
Modif 2	Dice's	69.91	56.49	100	76.02
	Jaccard	53.76	39.37	100	61.31
	Cosine	71.14	57.89	100	76.18
Modif 3	Dice's	86.82	67.22	76.02	100
	Jaccard	76.70	50.63	61.31	100
	Cosine	87.47	68.05	76.18	100

Table 2 shows the results of testing with test n-grams using value = 3. In this test, it can be seen that the similarity accuracy decrease compared table 1 using n=2. Original text compared with texts modified 1 and 2 give accuracy above 70%. While the Jaccard coefficient shows a drastic decrease in accuracy as the value of n-grams higher. Jaccard has a weakness when it comes to measure certain word [14]. Similarity value is influenced by the size of the small n value. The smaller the value of n-grams are used, the higher the accuracy of the similarity of the text file. This research has proven the opinion that similar research has been done previously [5][6][8] that the accuracy of the similarity is affected by the value of n-grams are used.

5 Conclusion

Based on the above research, ranging from the design stage, to the analysis of test results, model of distance-based similarity measure can be implemented successfully with Rabin-Karp. Cosine similarity show better performance in generating value accuracy with some scenario modified text compared to the Dice coefficient and Jaccard coefficient. It can achive above 90% high accuracy between Original text compared with modified text by change sentences position because similarity search based on the corresponding hash value. Simple cosine model is expected to be used as an alternative type of statistical algorithms that implement together with character n-grams as parameter in the process especially to detect plagiarism in Indonesian text.

The smaller the value of n-grams are used, the higher the accuracy of the similarity of the text file. Accuracy value only used as a reference to detect plagiarism, but we need more aspect to investigate it like writing style, citation uses, and others.

References

1. Zaka, B.: Theory and Applications of Similarity Detection Techniques. Graz University of Technology, Austria, Institute for Information Systems and Computer Media (IICM) (2009)
2. Stein, B., zu Eissen, S.M.: Near Similarity Search and Plagiarism Analysis. Presented at the 29th Annual Conference oF the German Classification Society (GfKI), Berlin (2006)
3. Wibowo, A., Sudarmadi, K., Barmawi, R.: Comparison between fingerprint and winnowing algorithm to detect plagiarism fraud on Bahasa Indonesia documents. In: Inf. Commun. Technol. ICoICT 2013 Int. Conf., pp. 128–133 (March 2013)
4. Karp, R.M., Rabin, M.O.: Efficient randomized pattern - matching algorithms. IBM J. Res. Dev.-Math. Comput. Arch. 31(2), 249–260 (1987)
5. Nugroho, E.: Perancangan Sistem Deteksi Plagiarisme Dokumen Teks dengan Menggunakan Algoritma Rabin Karp. Jurusan Ilmu Komputer, Universitas Muhammadiyah Malang (2011) (unpublished)
6. Salmuasih, A.S.: Implementasi Algoritma Rabin Karp untuk Pendeteksian Plagiat Dokumen Teks Menggunakan Konsep Similiarity. In: Semin. Nas. Apl. Teknol. Inf., SNATI 2013, pp. 23–28 (June 2013)
7. Kurniawati, A., Wulandary, L., Wicaksana, I.W.S.: The Comparation of Similarity Detection Method on Indonesian Language Document. In: 2nd Int. Conf. Soft Comput. Intell. Syst. Inf. Technol., pp. 285–289 (2010)
8. Surahman, A.M.: Perancangan Sistem Penentuan Similarity Kode Program Pada Bahasa C Dan Pascal Dengan Menggunakan Algoritma Rabin-Karp. J. Sist. Dan Teknol. Inf. 1(1) (2013)
9. Barnard, J.: Chemical Structure Representation and Search Systems. Presented at the Chemical Informatics Software & Consultancy Services, Sheffield, UK (November 18, 2003)
10. O'Connor, B.: F-scores, Dice, and Jaccard set similarity (April 11, 2012), http://brenocon.com/blog/2012/04/f-scores-dice-and-jaccard-set-similarity/ (accessed: March 03, 2014)
11. Approximate String Matching, http://cs.uef.fi/~zhao/Link/Similarity_strings.html (accessed: March 04, 2014)
12. Purwitasari, D., Kusmawan, P.Y., Yuhana, U.L.: Deteksi keberadaan Kalimat Sama sebagai Indikasi Penjiplakan dengan Algoritma Hashing berbasis N-Gram. Institut Teknologi Surabaya, Surabaya (2010)
13. Discussion of Similarity Metrics: Jaccard / Tanimoto Coefficient. Similarity Metrics, http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/sphilip/tani.html (accessed: March 04, 2014)
14. Niwattanakul, S., Sinthongchai, J., Naenudorn, E., Wanapu, S.: Using of Jaccard Coefficient for Keywords Similarity. Presented at the Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 (IMECS 2013), Hong Kong, vol. 1 (2013)

Document Searching Engine Using Term Similarity Vector Space Model on English and Indonesian Document

Andreas Handojo, Adi Wibowo, and Yovita Ria

Informatics Engineering Department
Faculty of Industrial Technology, Petra Christian University
Jl. Siwalankerto 121-131, Surabaya 60236, Indonesia
{handojo,adiw}@petra.ac.id

Abstract. In line with technology development, the number of digital documents increase significantly, this will make process to the search a particular documents experience a little problem. Therefore, the role of search engines is become inevitable.

Usually, search engines conduct a searching process simply by looking at the similarities between keywords (that inputed by user) and terms in a document. In this research, we try to implement Term Similarity Vector Space Model (TSVSM), a method that also saw the relationship between the terms in the document. The relationship between terms in a document is calculated based on the frequency of occurrence in a paragraph. So this will make possible to search documents that do not contain the exact keywords that inputed, but have terms that related to those keywords.

We also try to implement TSVSM to English language documents from CiteseerX journal collection [1]. In this research we also want try it to Indonesian language documents from journal collection on Petra Christian University Research Center (both in pdf format, with total 14.000 documents). This application was built using Microsoft Visual Basic.Net 2005 and PHP.

Based on testing, this method can establish relationships between related terms that can find documents that do not contain keywords but contains terms that relate to the keyword. Time that needed to search document in Indonesian language journal is relative longer than in English language journal.

Keywords: Term Similarity Vector Space Model, Term Co-Occurrence Vector, Term Co-Occurrence Matrix, Search Engine.

1 Introduction

With the development of technology, the data storage in physical form began to change into digital storage to simplify storage and retrieval at the future. To search a digital document that needed, we usually using a search engine that is generally looking terms in a document that have the same terms to the keyword that entered by the user (user queries). Because the search engines generally do not looking the possibility of similarity in terms in the document.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 165–173, 2015.
DOI: 10.1007/978-3-662-46742-8_15

For example, when a user searches for documents with the keyword "computer" then it will only display documents that have the keyword "computer". But there still are some documents that do not have the keyword "computer" but associated with the keyword "computer" such as "laptop", "notebook", "software", etc. This resulted will reduce recall in searching the relevant documents that given by keywords.

To overcome these problems, in this research we will try to use Term Similarity Vector Space Model (TSVSM) which have assumed that each term has a relationship to another term so that the search process is not only based on the keywords that are inputed by user, but also search the relationships between terms in the documents that stored the database. Relation or similarity between the terms is calculated based on the frequency of a term appears simultaneously with the other terms in a paragraph.

The results of these search engines will be display as a list of documents that relevant to the keyword, in order of its relevance. Through this method the recall of search results expected to be improved. This study is a continuation of our previous studies about search engine using Topic-Based Vector Space Model [2].

The parsing process will be done by reading the document (in pdf format) using PDFBox API library [3]. Then the stemming process will be done using Porter Stemming algorithm [4]. For the repository we used Oracle 10g database, while applications created using Microsoft Visual Basic .Net 2005 (application parsing and indexing) and Microsoft ASP .Net 2005 (application engines).

Like classical Vector Space Model method, the Term Similarity Vector Space Model also assumes that each document is represented a number of dimensions corresponding vector with the number of terms in the document. The difference is the assumption in the calculation of the relationship between terms, where a term is assumed to be a vector whose dimensions established by the terms of the other on the vector space [5]. So the dimensions of the term vector can be obtained through the creation of co-occurrence vector.

2 Vector Space Model

Vector Space Model is an algorithm that can be used to measure the similarity of documents with keywords or other documents which are represented in the form of vectors [6]. This model is based on the idea that the meaning of a document formed by the words that contained. One of the improvements on the vector space model is to expand the document vector and keyword vector. In the Vector Space Model, documents and keywords can be defined as follows [7]:

$$d_j = \left(w_{1,j}, w_{2,j}, w_{3,j}, \ldots, w_{t,j}\right) \quad (1)$$

$$q = \left(w_{1,q}, w_{2,q}, w_{3,q}, \ldots, w_{t,q}\right) \quad (2)$$

Each dimension that created document vector or keyword vector is determined by the existing of the terms. Usually the term is assumed as a single word, keyword, or phrase. If a word is considered as a term, then the number of dimensions from a document is the number of words that exist in the document. Several ways to calculate

the weight of each dimension is known as a term weighting that always have been developed. One way of term weighting is Term Frequency-Inverse Document Frequency (TF-IDF).

3 Term Similarity Vector Space Model

Similarity Term Vector Space Model is the development of the Vector Space Model which assumes that the terms on a document is also a vector which the dimensions established by the other terms contained in the vector space [6]. A term co-occurrence matrix is calculated directly from the document to establish basic based to calculating the term similarity [7]. The calculation of term co-occurrence matrix is seen from the frequency of occurrence of a term with another term in each paragraph. As shown in figure 1, where [7]:

1. Each row in this matrix represents the distribution of a term x_i.
2. Each column represents another distribution, y_j term that appears near term x_i.
3. Interchanges rows i and columns j, m_{ij} is the co-occurrence frequency between x_i and y_j is extracted from the document collection.

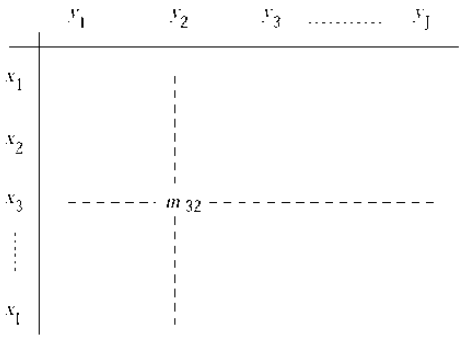


Fig. 1. Co-occurrence Matrix

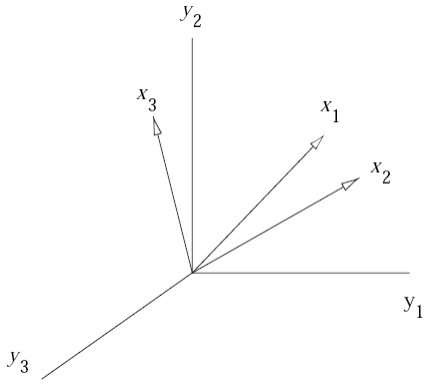


Fig. 2. Co-Occurrence Matrix Representation in Vector

Each dimension of this vector is associated with the word yj that represent the column matrix. In figure 2 is described as third-term representation of the first vector x1, x2, and x3 in figure 1 in a dimension that associated with the words y1, y2, and y3 from the collection of documents [7].

Similarity between terms can be seen from the large of cosine angle which is owned by the two term vectors that be compared. If the two vectors that term have a high similarity, the two terms are considered to have a relationship [7]. The formula for similarity searching process as follows:

$$sim(\vec{a}, \vec{b}) = \cos\theta = \frac{\sum_{i=1}^{N} W_{i,a} x W_{i,b}}{\sqrt{\sum_{1=1}^{N} W^2{}_{i,a}} x \sqrt{\sum_{1=1}^{N} W^2{}_{i,b}}} \quad (3)$$

Description:

$\vec{a}$: Vector term a

$\vec{b}$: Vector term b

$W_{i,a}$: Dimension of the i term in the vector-a

$W_{i,b}$: Dimension of the i term in the vector-b

N : Number dimensions of the vector term

Similarity searches have been done after finding the term vector of the co-occurrence matrix. After the words, documents, and keywords are represented as vectors in the same vector space, the keyword vectors will be compared with each vector of the document. The closest document vector to the vector keyword in the space will be delivered to users as an answer. The closeness of each vector viewed from the angle cosine between the documents with the formula as follows:

$$sim(dj, q) = \cos\theta = \frac{\sum_{i=1}^{N} W_{i,j} x W_{i,q}}{\sqrt{\sum_{1=1}^{N} W^2{}_{i,j}} x \sqrt{\sum_{1=1}^{N} W^2{}_{i,q}}} \quad (4)$$

Description:

dj : vector of the document j

q : vector of keyword

wi,j : i dimension of document vectors j

wi,q : i dimension of the keyword vector

4 Precision and Recall

Precision and recall is one of the methods that used to measure the performance from an information retrieval system. These measurements will compare between the relevant documents that obtained from the search results and the total number of documents [8]. This precision will be calculation using the following formula:

$$\text{Precision} = \frac{relevant\ document\ retrieved}{document\ retrieved} \quad (5)$$

Recall is number of relevant documents that obtained from search results divided by the total number of relevant document contained in the entire collection of documents.

$$\text{Re}call = \frac{relevant\ document\ retrieved}{total\ relevant\ document} \tag{6}$$

Ideally, recall and precision should be worth one, mean that the system gives the results of all relevant documents without showing any irrelevant. Unfortunately, in reality this is almost impossible to achieve.

5 System Design and Implementation

Design system flowchart from Document Searching Engine application can be shown on Figure 3.

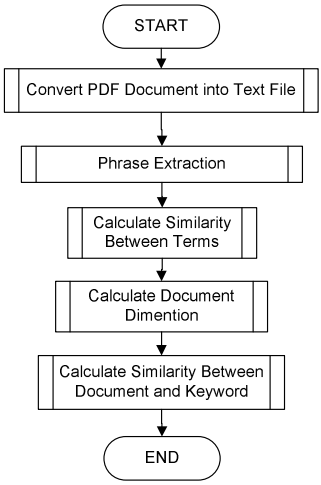


Fig. 3. Flowchart Searching Engine System

Description:

1. Convert PDF Document into Text File

In this process PDFBox library is used to convert pdf files into text files. This library is a library from open source Java programming language. This library will be running on the programming language that used for this search engine (Microsoft Visual Studio .Net 2005 and IKVM .Net libraries [9]).

2. Phrase Extraction

After changing pdf documents into text, then the next process is phrase extraction. This process is to find a set of words that have more meaning that can be seen from the frequency of collection of these words that appear simultaneously. These calculations is done by taking each n-word in each sentence in sequence and find word that occur together more often as word that have a deeper meaning.

3. Calculate Similarity between Terms

This process is a major part from this search engine. The calculation of similarity is to find the relation between terms. The determination of similarity is based on the closeness of the vector of a term with another term in a vector space. The dimensions of this term is determined from the frequency of a term appears together with other terms. To speed up the process and improve the outcome, this process required term removals that have too little relationships with the other terms.

4. Calculate Document Dimension

The calculation of each dimension of documents is done by multiplying the term frequency (TF), inverse document frequency (IDF), and term vectors.

5. Calculate Similarity between Document and Keyword

In this process, every document and the keyword is considered a vector in vector space and the similarity will be search by comparing the vector space.

6 System Implementation and Testing

We try to implement this search engine using research document in English and Indonesian language (in pdf format), but this system also can process documents in txt. First enter the document into the system (Figure 4).

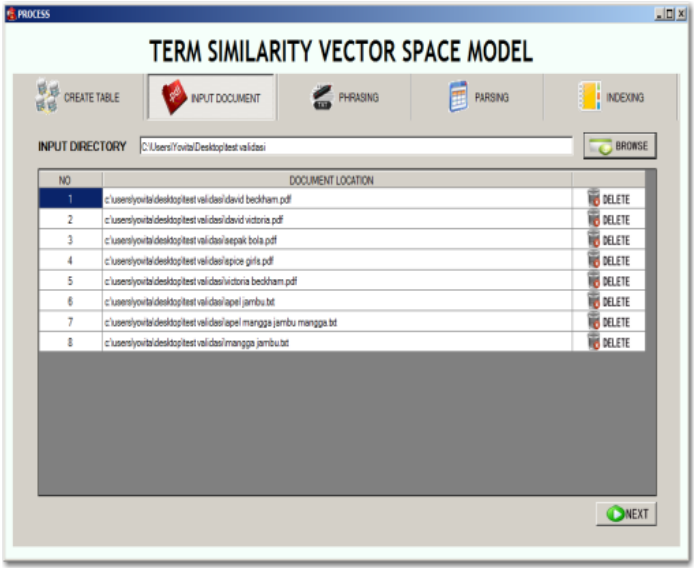


Fig. 4. Document Input Page

Once the documents are inserted, the process of phrase extraction, parsing, and indexing can be done (Figure 5), this will consume a little time. After that searching process is done by entering the keyword that user looking for (Figure 6), the search results will display the documents that have a correlation (on the right side) which also has a link to the document and the terms that have a correlation (on the left side).

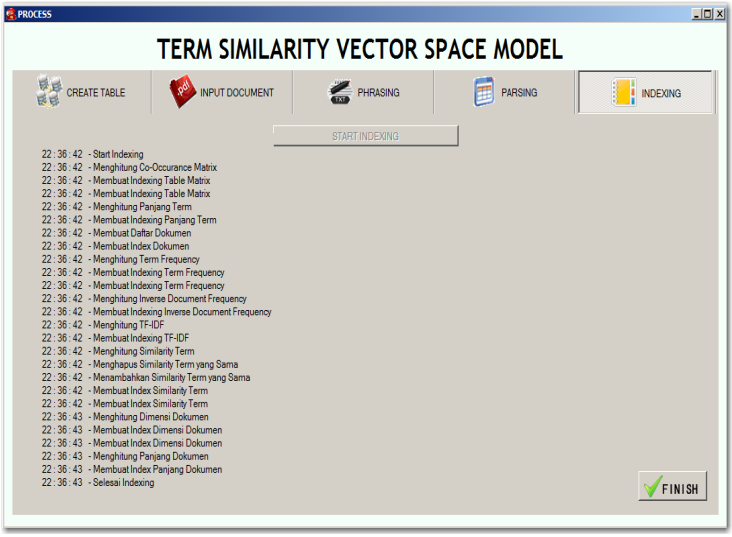


Fig. 5. Indexing Page

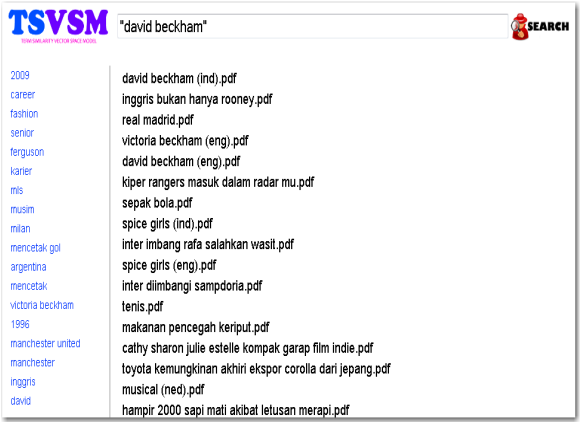


Fig. 6. Process Result Page

The following figures are the results from our experiment on TVSM.

1. Phrase Extraction

As seen on figure 7, the increasing number of terms versus length of time on phrase extraction O(0.28n+71). While the increasing size of the file versus time O(46.63n+40.19).

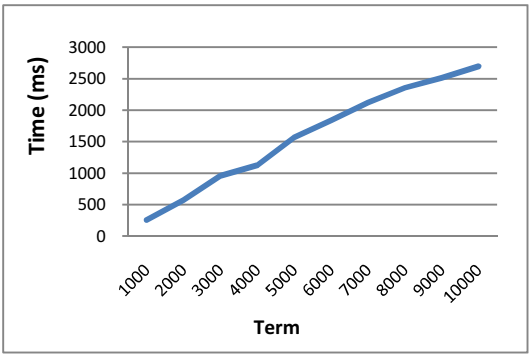


Fig. 7. Amount of Term vs Time on Phrase Extraction

2. Parsing

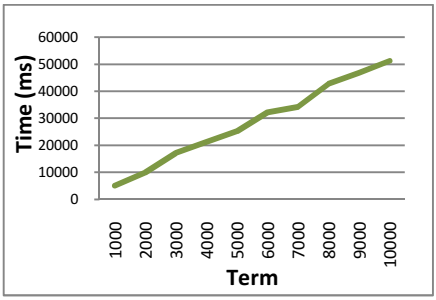


Fig. 8. Amount of Term vs Time on Parsing

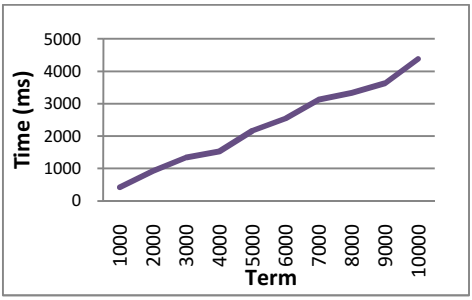


Fig. 9. Amount of Term vs Time on Indexing

As seen on figure 8, the increasing number of terms versus parsing time O(5.15n+225). While the increasing size of the file versus time O(864.97n-403).

3. Indexing

As seen on figure 9, the addition amount of term versus indexing time O(0.42n+15.5). While the increasing size of the file versus time O(70.87n-30.535).

The increasing number of terms, the more time that required on searching document O(0.11n-63). As shown on figure 10. While the increasing size of the file versus time O(18.04n-74.2). While the results of tests on recall and precision of 7 keywords on the 100 documents (figure 11). It appears that search engines have a maximum recall and a reasonable precision.

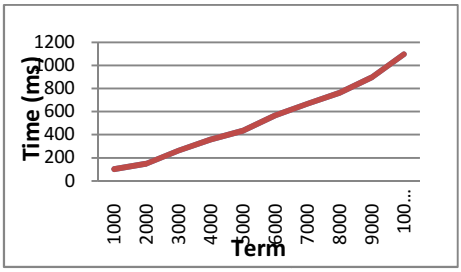


Fig. 10. Amount of Term vs Time on Searching

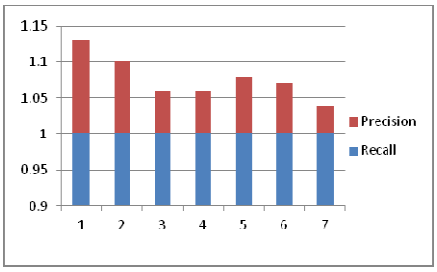


Fig. 11. Precision vs Recall

7 Conclusion

Based on the testing results that performed on the system it can be concluded that Term Similarity Vector Space Model can provide to produce documents that not contain the keyword but contains other keywords that are related. The Indexing process consume a lot of time, because this algorithm will be compares the relationship between terms based on term occurrences in a paragraph. Based on testing of the searching process, the more terms there are, the more time it takes to increase linearly. The searching process for document in Indonesian language is only takes a little bit more time than English language document, especially on the stemming process, but not too significant.

References

1. Citeseerx, http://citeseerx.ist.psu.edu/index (retrieved September 20, 2014)
2. Wibowo, A., Handojo, A., Halim, A.: Application of Topic Based Vector Space Model with WordNet. In: International Conference on Uncertainty Reasoning and Knowledge Engineering (2011)
3. The Apache PDFBox. A Java PDF Library, https://pdfbox.apache.org/ (retrieved October 2, 2014)
4. Karaa, W.B.A., Gribâa, N.: Information Retrieval with Porter Stemmer: A New Version for English. In: Nagamalai, D., Kumar, A., Annamalai, A. (eds.) Adv. in Comput. Sci., Eng. & Inf. Technol. AISC, vol. 225, pp. 243–254. Springer, Heidelberg (2013)
5. Kikui, G.: Term-list translation using mono-lingual word co-occurrence vector. NTT Information and Communication System Labs, Japan (1998)
6. Baeza, Y.R., Frakes, W.B.: Information retrieval data structures and algorithms. Prentice Hall PTR, Upper Saddle River (1992)
7. Grossman, D., Freider, O.: Information retrieval: Algorithms and heuristics, 2nd edn. Springer, Netherland (2004)
8. Davis, J., Goadrich, M.: The Relationship Between Precision-Recall and ROC Curves. In: The 23rd International Conference on Machine Learning, Pittsburgh, PA (2006)
9. Ikvm .net home page. Ikvm .net, http://www.ikvm.net/ (retrieved September 20, 2014)

Knowledge Representation for Image Feature Extraction

Nyoman Karna[1], Iping Suwardi[2], and Nur Maulidevi[2]

[1]School of Electrical Engineering, Telkom University, Bandung, Indonesia
aditya@telkomuniversity.ac.id
[2]School of Informatics and Electrical Engineering, Institut Teknologi Bandung, Bandung, Indonesia
{iping,ulfa}@stei.itb.ac.id

Abstract. In computer vision, the feature(s) of an image can be extracted as information using deep learning approach. This type of information can be used for further processing, for example to establish a visual semantic, which is a sentence that gives a description about the image. Usually this type of information is stored in a database point of view, which explains the relation between image feature and image description. This research proposes knowledge representation point of view to store the information gathered from image feature extraction, which in return, some new benefits can be obtained by using this approach. Two benefits that can be delivered by using knowledge representation instead of database point of view are integration capability with another source of information from related knowledge-based system and possibility to produce a high-level specific knowledge.

Keywords: Representation, knowledge model, image feature extraction, image description, semantic network.

1 Introduction

Phrase "a picture worth a thousand words" has been used for decades in computer graphic, which explains that one picture can be described in one thousand words. In image processing area, right now it is a very common and trend research on identifying an image and creates a sentence that can roughly describe what it is in the picture. The information extraction process from image to sentence is called visual semantic. Based on ACM classification codes, this research area is part of artificial intelligence. Where artificial intelligence coded as I.2, then visual semantic is coded as I.2.10 (Vision and Scene Understanding) along with the Knowledge Representation Formalisms and Methods, which is coded as I.2.4.

Previously it was only a dream to have a computer vision that can actually understand image by creating a sentence that provides the description of the image. But in the last 10 years the field of computer vision has made considerable progress. Described in Figure 1, the process of visual semantic itself consists of 3 steps. The first step is to extract locally measured features from an image. The features are invariant descriptors that cancel out accidental circumstances of the recording caused by differences in lighting, viewpoint, or scale. The second step is to project the descriptors per

R. Intan et al. (Eds): ICSIIT 2015, CCIS 516, pp. 174–182, 2015.
DOI: 10.1007/978-3-662-46742-8_16

pixel onto one of thousands of words. These are summarizations of one local patch of the image describing a single detail: a corner, texture, or point. In the third step, a machine-learning algorithm converts the visual words into one of the semantic concepts. In fact, it assigns a probability to all of the concepts simultaneously. These probabilities are used to rank images in terms of concept presence [1].

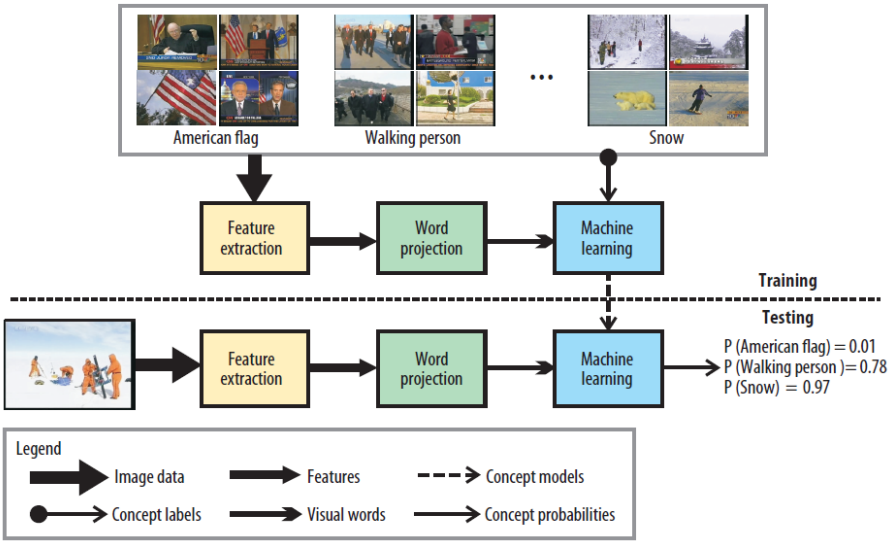


Fig. 1. General scheme for detecting visual concepts in images (Source: [1])

The visual semantic of the image, which creates feature extraction, can be described by using several methods, but still the output has the same elements, which is a vector. Visual semantic extracts some features from the image to identify objects in the image. This process produces a mid-level of semantic concepts. But human, on the other hand, can create a high-level specific semantic concept, based on their knowledge and emotion. The problem with feature extraction is the output is still in mid-level semantic concepts, where human can extract the concept and also relating them with other knowledge and experience (http://www.research.att.com/projects/Video/VisualSemantics/). This is explained in Figure 2. The gap between machine output and human output is called semantic gap. Shortening the gap is the problem that we would like to rise into our attention by proposing an introduction of knowledge model to be utilized as a media to store and manage the information obtained from image feature extraction.

If we want to shorten the gap between machine extracted information (mid-level semantic concepts) and human observed information (high-level information), we believe it will be prudent to insert a model that has the ability to store the machine extracted information and be used to provide relationship with other knowledge gathered from another source. This research proposes an intermediary process in the machine learning, where a knowledge model is introduced into the machine learning. Proposing knowledge model within the machine learning is to ensure that a high-level specific can be attained by using the information stored in the model.

Fig. 2. Semantic gap between machine's extracted information and human's observed information (Source: http://www.research.att.com/projects/Video/VisualSemantics)

2 Related Works

Knowledge is a proposition that contains facts and is definitive, while representation is a relation between two domains, the symbol and what the symbol represents. Knowledge representation is a field of study that explores symbol formalism, which is used to represent knowledge [2,3]. Knowledge itself has many interpretations and so far there have been no agreed definitions for knowledge. Although there are no definitive definitions for knowledge, we can look at several aspects, machine learning, expert systems, and knowledge management. In machine learning, information is acquired and retained (stored) for future recall to obtain knowledge from the existing information. In an expert system, information is acquired from an expert's knowledge and retained for future recall to obtain the expert's knowledge. In knowledge management, a large amount of knowledge is also basically stored in information fact types, which will be easily recalled to expand the user's knowledge.

Every knowledge representation needs a specific language to provide an optimal way to represent a symbol. The symbol itself is needed to represent the knowledge in such a way that it can be easily understood by a simple machine, that is, a program. Language will provide an effective way to represent knowledge using three aspects [2], which are as follows.

1. Syntax, to define how the formalism of knowledge representation forms a sentence with clear and standard structure by building it with logical symbols (punctuation, connectives, variables) and non-logical symbols (function and predicates)
2. Semantic, to define how the formalism of knowledge representation forms a sentence with a structure that can be understood through:
   a. interpretation
   b. denotation
   c. satisfaction
3. Pragmatic, to define how the formalism of knowledge representation forms a meaningful sentence.

Knowledge representation is a branch of artificial intelligence, which is a field of study that explores how to represent information that is acquired from anywhere in a

format that can be understood by a program. A program here refers to a knowledge-based System, such as a machine learning system or an expert system. The knowledge representation itself can also be described as a knowledge model, because it can explain the model of the knowledge, the syntax and the semantics of the information. When we choose the model of knowledge it depends on three aspects [4].

1. Problem, what is the problem to be addressed?
2. Goal, what should the knowledge fulfill?
3. Proposed solution, how does the knowledge solve the problem?

Knowledge model is used to represent knowledge that is declarative with specific and definitive meaning. A declarative sentence that fulfills all 3 aspects (syntax, semantic, and pragmatic) can be used to establish knowledge representation using several formalism [2], [4]:

1. Semantic Network, such as ontology, IS-A link, and On Concept Algebra, concentrates on expressing the taxonomic structure of categories of objects and the relations between them
2. Frame, to store knowledge acquisition from NLP (Natural Language Processing), focuses on parsing and store key point information
3. Production System, such as Rule-based (IF-THEN), reflects the notion of consequence
4. Logic-based, such as First-Order Logic and fuzzy logic, enhances the detail expressiveness of other knowledge model (Semantic Network, Frame based, and Production System)

Logical foundations for interpreting semantic networks are to create a clear distinction between the intensional logic of terminological definitions, and the extensional logic of asserting properties of individuals. This distinction led to the separation of the 'T-box' and 'A-box' in later implementations of representation and inference systems deriving from semantic networks. It became accepted in the 1980s that semantic networks should be treated as a visual language for some subset of FOL without any extra logical inference capabilities, and that inference systems based on them should be assessed in terms of soundness and completeness in the same way as any other proof system for a formal logic [5]. This foundation provides a basic background on why this research propose semantic network as the knowledge model to represent the feature information extracted from image.

When an image is processed to extract the features in the image, the output of the information is a series a variable that contain a specific numerical value. This series of numerical value is called vector, which will be processed further to find similarity or distance between vectors. The similarity of several vectors will define the feature characteristic of the image [6,7], thus the relation between a specific image feature and the distance vector should be stored and maintained for further process by artificial intelligence.

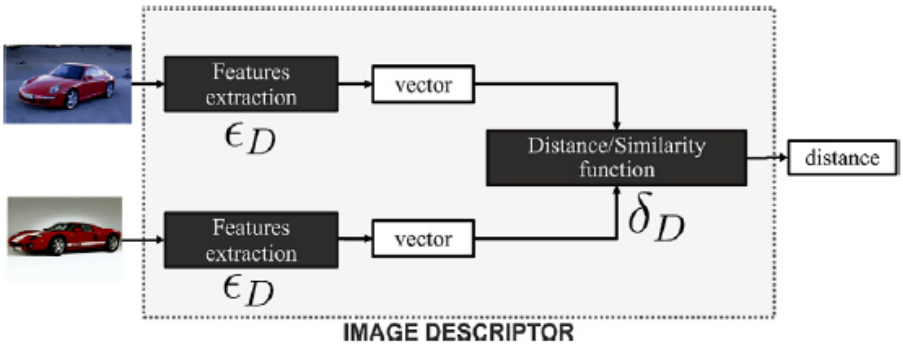


Fig. 3. Image descriptor that shows output of the feature extraction is vector [7]

This research proposes a knowledge model to represent the relationship between a specific image feature and the vector that describe the distance between sample vectors. This relationship will be stored in knowledge model, especially in semantic network representation, to be utilized further for enhancing the visual semantic. The utilization of knowledge model in this process is to gain benefit for integration with machine learning. This integration can further provide high-level visual semantic based on another source of information.

By providing knowledge in multiple knowledge-based systems, we can create knowledge transfer, knowledge integration, and knowledge sharing. The knowledge transfer capability ensures all the knowledge within one system can be transferred to a new system. The knowledge sharing capability ensures the knowledge within one system can be shared to enhance the knowledge of another system. The knowledge integration capability ensures the knowledge within one system can be used by another system [4]. This research will use knowledge integration between knowledge-based systems to create new knowledge. The capabilities are shown in Figure 4.

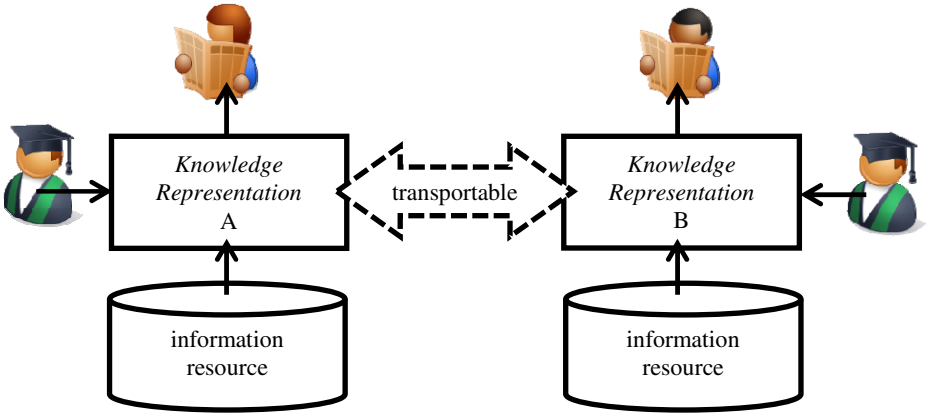


Fig. 4. Knowledge Transfer, Sharing, and Integration between 2 systems [4]

3 Proposed Solution

The gap between mid-level semantic concept and high-level specific semantic concept can be reduced if the mid-level is stored in a specific knowledge model. The information stored in the knowledge model can be further collaborated with the information obtained from another source to create high-level specific knowledge. By combining the mid-level semantic concept with information from another knowledge-based system, it will provide a mechanism to obtain high-level specific knowledge. This mechanism is explained in Figure 5.

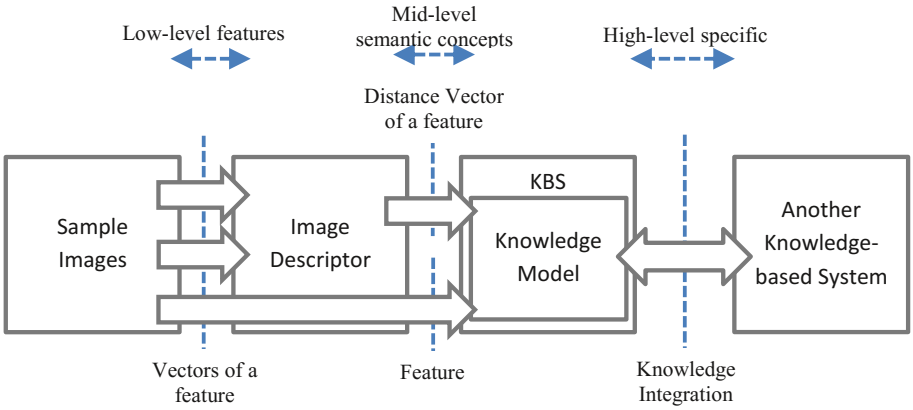


Fig. 5. Knowledge integration to obtain high-level specific knowledge

From sample images we can extract low-level features, which are vectors of a feature that highly related to the image itself. These vectors naturally will be processed by Image Descriptor to obtain a distance vector that we would like to relate to a specific feature. The relation between distance vector and the feature it represent is the mid-level semantic concept, which is highly related with only the given sample images. To achieve the high-level specific knowledge, we need to store the mid-level semantic concepts in a specific knowledge model within a KBS (knowledge-based system) and integrate the knowledge within with another source of information from another KBS to obtain a high-level specific knowledge.

The knowledge model we propose in this research is the semantic network, that provides knowledge representation in a visual network with nodes (knowledge atoms) and vector labeled interconnection that interconnect between nodes [4]. The reason why we choose the semantic network for the knowledge model to represent the relation between distance vector and the feature it represent is because semantic network has good representation for integration with another knowledge-based system because of the interconnection that can easily be connected with another semantic network from another knowledge-based system.

We propose semantic network knowledge model to store mid-level semantic concept. Because the mid-level semantic concept is in the format like Formula 1 where the knowledge atom is a relationship between distance vector and the feature it represented, we design the knowledge model to consist of nodes where each node

provide a storage to place the relationship and interconnection where each interconnection provide an explanatory of the relation between nodes. An example of the visualization of semantic network is drawn in Figure 6. In this figure each node contains one KA or knowledge atom (a relationship between a vector and a feature). Every node has a profile that describes the attribute of the knowledge atom, for example noun, object, verb, etc. Each of the node has a minimal one relation network with another node that describe the similarity with another node that share similar profile that use Levenshtein function described in Formula 2.

$$\text{Vector} : \{a_1, a_2, a_3, \ldots, an) \leftrightarrow \text{feature} \tag{1}$$

$$\text{Similarity}(KA_A, KA_B) = lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{(ai \neq bj)} \end{cases} & \text{otherwise.} \end{cases} \tag{2}$$

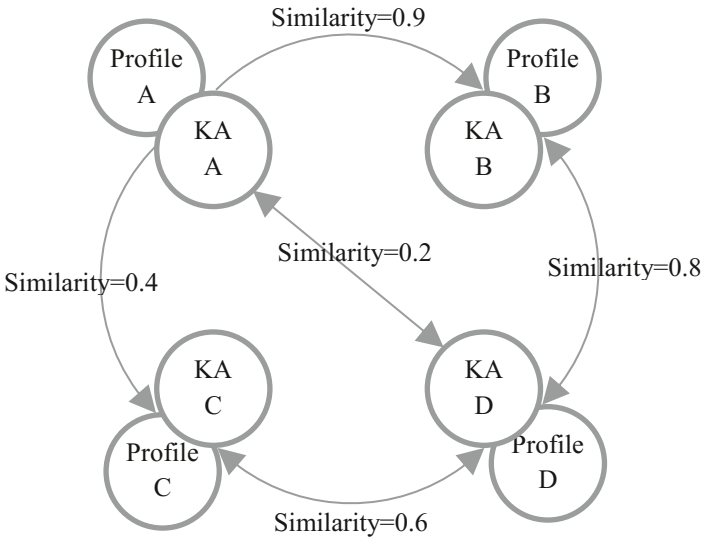


Fig. 6. Proposed semantic network knowledge model to store the mid-level semantic concept

4 Implementation Design

The implementation focuses on the establishment of knowledge-based system (KBS) to provide intermediary storage and processing to produce high-level specific knowledge. The mid-level semantic concepts is not the focus since it has already done by recent research [8] where more than 21,000 sample images are used as a data set to

produce the mid-level semantic concept data. In this research, the image feature from this data set was extracted using java library JFeatureLib (https://code.google.com/p/jfeaturelib), which in return provide data vector that will be stored in KBS. For representing the knowledge of the image feature, this research uses Protégé a tool provided freely to implement a semantic network (http://protege.stanford.edu). The "individuals" in Protégé represents the distance vector or mid-level semantic concept where the "class" represents profile of the vector. The "properties" in Protégé represents the similarity between profiles of the vector that will be used to help choosing the appropriate KBS.

To provide integration between KBS, this research uses knowledge query and manipulation language (KQML) [9]. This language provides the ability to retrieve information from a specific KBS and uses it to enhance a conclusion from several KBS, which in this case is to provide high-level specific knowledge. KQML provide two mechanisms beneficiary in this research, which are how to find which KBS has the data (mid-level semantic concept) needed and how to initiate and exchange the data. Another major problem that has to be resolve in this research is how to create a high-level specific knowledge from several mid-level semantic concepts obtained from each KBS.

5 Proposed Benefits

By providing a mediatory for storing the mid-level semantic concept using a semantic network knowledge model within knowledge-based system, we can obtain the high-level specific knowledge by integrating with another source of information. This source of information can be found from another related knowledge-based system that can supply additional information as premises to obtain a new and more high-level specific knowledge. By integrating with several knowledge-based systems, we can even obtain much more high-level specific knowledge. This is the reason why we propose semantic network as the knowledge model to represent the mid-level semantic concept, because with semantic network we can easily interconnect a knowledge-based system with another.

6 Conclusion and Future Works

An image descriptor can only output a vector that represent a feature from images. The extracted features are highly related with the content of the image itself, which is why we identified them as mid-level semantic concept. There is no possibility for image descriptor to provide a high-level specific knowledge. This semantic gap is the problem we want to resolve. Our research is to propose a knowledge model that store the vector in semantic network and use them with another source of information to obtain a new and high-level specific knowledge.

As a future work, we are going to implement the semantic network using Protégé and KQML for integration and use them to overcome the semantic gap by creating a high-level specific knowledge.

References

1. Snoek, C.G.M., Smeulders, A.W.M.: Visual-Concept Search Solved? Computer 43(6), 76–78 (2010)
2. Brachman, R.J., Levesque, H.J.: Knowledge Representation and Reasoning. Morgan Kaufmann, San Francisco (2004)
3. Sowa, J.F.: Knowledge Representation: Logical, Philosophical, and Computational Foundations. Course Technology, CENGAGE Learning, Boston (2000)
4. Karna, N., Supriana, I., Maulidevi, U.: Intelligent Interface for a Knowledge-based System. Telkomnika Telecommunication, Computing, Electronics and Control 12(4), 1096–1104 (2014)
5. Gaines, B.R.: Designing Visual Languages for Description Logics. Journal of Logic, Language and Information 18(2), 217–250 (2009)
6. Salinas, R.M., Aguirre, E., Silvente, M.G.: Detection of doors using a genetic visual fuzzy system for mobile robots. Autonomous Robots 21(2), 123–141 (2006)
7. Penatti, O.A.B., Valle, E., Torres, R.S.: Comparative study of global color and texture descriptors for web image retrieval. Journal of Visual Communication and Image Representation 23(2), 359–380 (2012)
8. Rachmawati, E., Supriana, I., Khodra, M.L.: Multiclass Fruit Classification of RGB-D Images using Color and Texture Feature (in press)
9. Finin, T., Fritzson, R., McKay, D., McEntire, R.: KQML as an agent communication language. In: CIKM 1994: Proceedings of the Third International Conference on Information and Knowledge Management, pp. 456–463. ACM (1994)

Using Semantic Similarity for Identifying Relevant Page Numbers for Indexed Term of Textual Book

Daniel Siahaan[1] and Sherly Christina[2]

[1]Dept of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
daniel@if.its.ac.id
[2]Dept. Informatics Universitas Palangkaraya, Palangkaraya, Indonesia
ayironaldo@gmail.com

Abstract. Back-of-book index page is one of navigation tools for reader. It helps reader to immediately jump to a page that contains relevant information regarding a specific term. It helps reader to retrieve information about specific topics in mind without having to read the complete book. Indexed terms are usually determined by author based on one's subjective preference on what indications should be used to decide whether a term should be indexed and what pages are relevant. Therefore, indexing a book inherits subjectivity of author side. The book size is proportional to the indexing effort and consistency. This leads to the fact that page numbers are not always referred to relevant pages. This paper proposes an approach to identify relevancy of a page that contains an indexed term. This approach measures the semantic relation between indexed term with the respective sentence in the page. To measure the semantic relation, the approach utilizes semantic distance algorithm that based on Wordnet thesaurus. We measure the reliability of our system by measuring its degree of agreement with the book indexer using kappa statistics. The experimental result shows that the proposed approach are considered as good as the domain expert, given average kappa value 0.6034.

Keywords: book indexing, back-of-book index, relevant page number, semantic relation.

1 Introduction

Back-of-book index has been part of book publishing for decades. Reader uses back-of-book index as a mean to instantly jump to a specific page(s) suggested to contain respective indexed term. This facility is suitable to satisfy the need of a reader who would like to retrieve relevant information about specific topic in mind without having to traverse through the book.

Set of indexed terms should represent topics discussed within a respective book [1]. An indexed term may take a form of a word, phrase, or alphanumerical text. Let's consider following example which illustrates a snapshot of a back-of-book index.

```
Software   11
Software Engineering     11,42
Software Requirements   77
```


R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 183–192, 2015.
DOI: 10.1007/978-3-662-46742-8_17

From the example we can see that a back-of-book index contains a list of indexed terms. Each indexed term indicates a term which is considered important and relevant to the book and a list of page numbers. Each page number of a term refers to a start of a sequence of pages contain the respective term. The amount of indexed terms grows relatively proportional to the number of book pages.

Author of a book usually selects terms to be indexed by analyzing documents manually based on one's expertise. The author bases the selection on one's subjective judgment. The author judges whether a term resides in the book is highly important and relevant with the main topic of the book or not. If it is then the term is listed in the back-of-book index. Furthermore, the author also need to judge whether each page that contain the indexed term indeed focusing the discussion or description on the respective indexed term or on other term.

Many sound proof algorithms available for extracting topics out of a text document. There are a number of works related to book indexing. Lahtinen [2]extracted indexed terms from document by combining the frequency of word appeared in the document and the result of linguistic analysis. Duque and Lobin [3] used natural language processing and ontology to find indexed term of a document. Medelyan and Witten [4] introduced KEA++, a machine learning for extracting terms. The algorithm is trained using indexed term candidate features which are originated from tf*idf weight and position of the indexed term first occurrance. To our knowledge, there is no previous work has been done in addressing the second issue of book indexing, i.e. identifying which page number that refers to a relevant page. Modern word processors, such as Ms. Word and TeXMaker, nowadays provide a mechanism to create back-of-book index from a book that is being composed. But author still needs to tag a mark on a candidate indexed term manually in order to be appeared in the index list. When the amount of book pages is larger, then the author consistency in judging whether a page is relevant or not is decreasing and the time consumed for this effort is extremely long. Therefore, our work is proposing an approach for identifying relevant page numbers for indexed term of a textual book. This approach can be used to help creating book indexed automatically. It is also as reliable as the human expert.

2 Indexed Terms and Page Relevancy

Inaccuracy and inconsistency in book indexing may be indirectly caused by the fact that indexing is relatively an expensive task. The list of indexed terms is subjective to perspective of the indexer (usually the author) on what terms are important and relevant to the main topic and which pages are relevant to each indexed terms. Selecting indexed terms and its relevant page numbers require expertise of the indexer on the context of the book and also domain of the main topic of the book.

Two mistakes that commonly found in book indexing, especially determining relevant page numbers, are indexed terms are not found in the referred page and unrecognized relevancy between the text in the referred page with the respective indexed term [1]. Let's consider the following example from a book entitled "Working with objects The Ooram Software Engineering Method" written by Reenskaug et al. (2001). Considering a term "type" as one of the indexed terms which appears in page number 27 and 370.

```
The type is the "what" abstraction. (page 27)
```

"Investment decisions are currently dominated by people who alternate between both types of activities, who understand the needs of production and the potential for improvement in the reusables." (page 370)

Both texts, in this case is two sentences, in page 27 and 370 contains the term "type." The indexed term "type" in the book refers to an abstraction of set of objects that have the same functionalities and attributes. The term "type" in page 27 indeed relevant to the context of indexed term "type" in the book. On the other hand, the term "type" in page 370 is not relevant to the context of indexed term "type" in the book. The sentence does not focus on the term "type".

3 Relevant Page Number Identifier

Our approach is divided into three main processes. The first process is sentence preprocessing. It converts a document element into a list of base-form words. The document element is selected from a referred page because it contains the indexed terms. The selected document element is then considered as a selected sentence. It starts by tokenizing the sentence to produce a list of tokens. Then the process removes any stop-word token. After all stop-word tokens are removed, the process reduces any inflected or derived part of word in a token and returns its base form. Consider the previous examples of indexed term (t) "type" and sentence s1 and s2.

```
t           :   type
s1 (p. 16)  :   The type is the "what" abstraction.
s2 (p. 43)  :   Investment decisions are currently
                dominated by people who alternate between
                both types of activities, who understand
                the needs of production and the potential
                for improvement in the reusables."
```

The sentence preprocessing process converts each sentence (s1 and s2) into a list of tokens. Any token that contains stop-word is removed. Each non-stop-word token is stemmed to get its base form. Given s1 and s2, the sentence preprocessing produce the following lists.

s_1 / t	type	abstraction
type	1	0.6

s_2 / t	investment	decision	current	dominate	people
type	0.43	0.43	0.11	0.11	0.67

alter-nate	type	activity	understand	need	production	potential	improvement	reusable
0.62	1	0.4	1	0	0.71	0.4	0.43	0.12

Fig. 1. Result of words similarity measurement of indexed term *t*

```
s1:  "type", "abstraction"
s2:  "investment", "decision", "current", "dominate",
     "people", "alternate", "type",
     "activity","understand","need","production","potenti
     al","improvement","resuable"
```

The second process is semantic relation measurement. It measures how close a sentence relates semantically to the context of the indexed term. It utilizes Wordnet thesaurus and uses fast-heuristics algorithm to measure the semantic relation between term and list of words produced from the first process [5]. The process first measures word relevancy between the indexed term with each of tokens of a sentence. The words relevancy measurement is represented as matrix M with the size of m×n, where m is the number of words in t and n is the number of tokens from the given sentence (s1 or s2). The value of Mij represents the degree of relevancy between ti and tokenj, in the scale of 0 to 1. If both words are found in Wordnet, the word relevancy measurement uses Wu and Palmer semantic similarity measurement formula [6]. If any of the words is not found in Wordnet, then word relevancy measurement uses Levensthein Distance Similarity formula [7]. Consider the aformention example, with term t and two lists of tokens (s1 and s2) . Fig. 1 shows the result of measuring the word similarity between indexed term t and each word in each of sentences.

Given the word relevancy measurements of matrix M, then the process calculates the degree of sentence relevancy of a given sentence. The sentence relevancy measurement is using a simple fast-heuristic [5]. Given the aforemention example, we can calculate the sentence relevancy of sentence s1 and s2 respectively as:

$$\text{Relevancy}(t,s\_i) = (\sum \max\{\text{value col\_i}\} + \sum \max\{\text{value row\_j}\})/(i+j) \quad (1)$$

```
Relevancy (t,s_1 )=(1+1.6)/(1+2)=0.8667
Relevancy (t,s_2 )=(1+2.97)/(1+14)=0.2647
```

The last process in relevant page number identification is the pigeon hole process. This process classifies a page number as relevant or irrelevant based on its sentence relevancy value. Given a contant threshold r-value, a page number that has sentence relevancy higher than r (Relevancy(t,si) > r) is considered a relevant page number, while a page number that has sentence relevancy equal or lower than r (Relevancy(t,si) ≤ r) is considered irrelevant page number.

The value of r is determined by using trial and error method. The initial value of r is set as 0.5. In the second iterations we increased and decreased the value by 0.1 repeatedly. Based on the most reliable result (indicated by the kappa value [8]), we decided the new r-value. In the third iterations, we increased and drecreased the value by 0,01 repeatedly and select the new r-value. We repeated this step until we found the optimum r-value.

4 Result and Discussion

4.1 Testing Data Set

For our experimentation, we composed a testing data set. The testing data set is used to measure the performance of our approach. The data sets are composed from 31 e-books that contain back-of-book index pages. The books are chosen from eight different disciplines. For each e-book we randomly selected a number of indexed terms. For each indexed term we prepared a relevant and irrelevant pages based on its back-of-book indexes. For the relevant page, a domain expert selected the most relevant sentence (and contains the indexed term) in the page referred by respective page number. For the irrelevant page, a domain expert also selected a sentence (that

Table 1. Experimentation Result

Domain	Number of Books	Number of Terms
Computer Science	6	125
Economy	3	56
Social and Politics	6	210
Education	3	29
Biology	3	108
Theology	3	50
Business	3	87
General	4	62
Total	**31**	**727**

contains indexed term) in the page not referred by respective page number. Both sentences should contain the respective indexed term. In total we have 1454 sentences,

which consist of 727 indexed terms. Table I shows the information about books used for composing our data set.

4.2 Experimentation Scenario

The experimentation scenario aims to identify the best threshold (r-value) given the degree of agreement between human expert and our proposed approach. This agreement is measured by kappa value [8]. The experimentation is conducted in two phases. In the first phase, we run the test over the data set ten times given the r-value range from 0.1 up to 1.0 with an interval of 0.1. The result is shown in Table II. We can see almost in all discipline the best threshold is in 0.5, except for books which is categorized as General. The highest and the lowest kappa values given threshold 0.5 are 0.78 for Theology books and 0.464 for General books. Table III shows average kappa value of each corresponding threshold. Notice that the threshold 0.5 produces the highest kappa value, i.e. 0.6034. The result suggests that the best threshold may

Table 2. The κ for Each disciplines: 1st phase

r	Comp Science	Economy	Social Politics	Education
0.1	0.000	0.000	0.000	0.000
0.2	0.000	0.000	0.000	0.000
0.3	0.016	0.000	0.004	0.000
0.4	0.096	0.214	0.104	0.172
0.5	**0.768**	**0.482**	**0.619**	**0.655**
0.6	0.632	0.089	0.247	0.310
0.7	0.384	0.035	0.119	0.103
0.8	0.200	0.035	0.080	0.103
0.9	0.128	0.000	0.071	0.103
1.0	0.112	0.000	0.066	0.103
	Biology	Theology	Business	General
0.1	0.000	0.00	0.000	0.000
0.2	0.000	0.00	0.000	0.000
0.3	-0.009	0.02	0.022	0.033
0.4	0.166	0.24	0.057	0.167
0.5	**0.518**	**0.78**	**0.494**	0.464
0.6	0.240	0.56	0.252	**0.516**
0.7	0.111	0.32	0.195	0.439
0.8	0.108	0.20	0.126	0.266
0.9	0.074	0.14	0.114	0.125
1.0	0.064	0.12	0.091	0.109

Table 3. the Average κ: 1st phase

r	κ
0.1	0.0000
0.2	0.0000
0.3	0.0097
0.4	0.1325
0.5	**0.6034**
0.6	0.3480
0.7	0.2067
0.8	0.1324
0.9	0.0918
1.0	0.0809

lie within 0.4 to 0.6 given the fact that the general pattern of kappa values seems to consistently increase from 0.4 to 0.5 and decrease from 0.5 to 0.6 given the highest point is in 0.5.

Table 4. The κ for Each disciplines: 2nd phase

r	Comp Science	Economy	Social Politics	Education
0.45	0.400	0.410	0.342	0.517
0.46	0.464	0.410	0.409	0.586
0.47	0.496	0.428	0.480	**0.689**
0.48	0.568	0.446	0.557	**0.689**
0.49	0.704	0.446	0.580	0.655
0.50	0.768	**0.482**	**0.619**	0.655
0.51	0.776	0.410	0.566	0.551
0.52	0.792	0.357	0.519	0.551
0.53	**0.802**	0.303	0.495	0.379
0.54	0.752	0.196	0.466	0.413
0.55	0.728	0.178	0.400	0.379
	Biology	Theology	Business	General
0.45	0.305	0.560	0.218	0.274
0.46	0.324	0.680	0.287	0.370
0.47	0.370	0.660	0.310	0.419
0.48	0.425	0.700	0.390	0.451
0.49	0.472	0.740	0.448	0.451
0.50	**0.518**	**0.780**	0.494	0.451
0.51	0.481	0.760	**0.505**	0.483
0.52	0.462	**0.780**	0.494	0.500
0.53	0.379	0.720	0.459	**0.516**
0.54	0.342	0.740	0.459	0.467
0.55	0.305	0.740	0.413	0.435

In the second phase, we repeated the experimentation by running the test over the same data set ten times given the r-value range from 0.45 to 0.55 with an interval of 0.01. As already mentioned, we repeated the experimentation to find the most reliable threshold, i.e. producing relatively highest kappa value. The result can be seen in Table IV. We can see that the best thresholds of each discipline are relatively varied. They are between 0.47 until 053. Table V shows average kappa value of each corresponding threshold. Notice that the threshold 0.5 still produces the highest kappa value, i.e. 0.6034. Based on this second phase result we can assume that threshold 0.5 produces the best kappa value. It can be used as threshold in determining page relevancy given an indexed term in back-of-book index. According to Landis and Koch [9] on interpretation of kappa value, we can see that our approach is considerably substantial in term of reliability compare to the domain expert judgment.

t \ s_1	asia	foremost	sell	chocolate	strawberry	milk	instead	skimmed	milk
foremost	0	1	0,12	0,11	0,1	0,12	0	0,12	0,12

t \ s_4	first	foremost	package	functional	sense	capable	protecting	product	minimum	cost
foremost	0	1	0	0,2	0,25	0	0,2	0,25	0,12	0,38

Fig. 2. Result of words similarity measurement of indexed term "*formost"*

It has average kappa value of 0.604, with 0.4611 and 0.7800 for the worst and best cases respectively, given the optimum threshold 0.543.

From the experimentation we also noticed that for a relevant page number (i.e. one that is listed in the back-of-book index) where our approach considers it as irrelevant, the sentence relevancy measurement is low. Let's consider the following sentences s3 and s4 and term t1 from "International Marketing" written by Sak Onkvisit and John J Saw (2007).

```
t           : foremost
s3 (p. 296): "In Asia, Foremost sells chocolate and
              strawberry milk instead of lowfat and
              skimmed milk."
s4 (p. 336): "First and foremost, a package must be
              functional in the sense that it is capable
              of protecting the product at minimum cost."
```

Sentence s3 is the sentence selected from the page number listed in the back-of-book index. Sentence s4 is the sentence selected from the page number not listed in the back-of-book index. Fig. 2 shows word similarity between indexed term "foremost" with each word in each sentences. The relevancy between indexed term "foremost" and sentence s3, i.e. Relevance(t, s3), is lower (0.269) than sentence s4 (0.34). Wordnet fails to recognize the relation between the indexed term "foremost" with words that form the sentence (except for the word "foremost", the maximum word relevancy value is 0.12). This is because the thesaurus is still limited and the work of recording the relation between words is still ongoing.

Another reason for the low relevancy measurement of relevant page may be related to the sentence length. Low relevancy tends to happen to long sentences. Consider the following two terms, i.e. "liquidation" and "subscriptions", from "Tax Deductions for Professionals" by A.S. Fishman (2008).

```
t1         :  liquidation
s5 (p. 47):  "Moreover, if the property is distributed to
              the shareholders on liquidation, they will
              have to pay tax on the amount that the
              property's fair market value exceeds the tax
              basis of their shares."
t2         :  subscriptions
s6 (p. 327): "License Fees, Dues, and Subscriptions."
```

T \ S	create	stored	procedure	perform	series	transact	command	server
stored	0	1	0,33	0,14	0,33	0	0,14	0,33
procedure	0,33	0,33	1	0,22	0,62	0,11	0,84	0,22

Fig. 3. Result of words similarity measurement of indexed term "*stored procedure"*

Given the indexed term t1, s5 is a relevant page with Relevance(t1, s5) 0.4058. The sentence s6 is a relevant page for indexed term t2. with Relevance(t1, s5) 1.0. The relevancy score of s5 is relatively lower compare to s5. Since word relevancy is A word has relatively few senses. This means that the higher semantic similarity of a word-sense pair, the less number of the pairs. The lower semantic similarity of a word-sense pair, the more number of the pairs. Therefore, long sentences tend to have more words with low similarity value compare to short sentences.

Although the experimentation result shows that the proposed approach is considerably reliable, the approach has not considered the fact that a term may comprise of more than one word, namely a phrase. The aggregation of relevancy between a phrase (in a token) with its indexed term may not represent the relevancy of the token with the indexed term. That is, the aggregation of relevancy between separate words of an indexed term (if itcomprise of more than one word) with a token may not represent the relevancy of the indexed term with the token. Let's consider the following example from a book entitled "Microsoft SQL Server 2000 – Database Design" written by Rozanne M. Whalen (2000).

```
t          :  stored procedure
s7 (p. 19):  "You can create a stored procedure to
             perform a series of Transact-SQL commands on
             your server."
```

During sentence processing, the term "stored procedure "is tokenized into separate tokens which contains each single word, i.e. "store" and "procedure". The term is later treated as separate token regardless its semantic meaning. Fig. 3 shows word similarity of indexed term "stored procedure" with each word in sentence s7. Since the term "stored procedure" is treated as two terms, i.e. "store" and "procedure", the degree of semantic similarity between the indexed term "stored procedure" with each token within s7 is different with the degree of semantic similarity between the term "store" or "procedure" with each token within s7.

The issue lies in the fact that Wordnet only provide word-sense pair but not phrase-to-word-sense or phrase-to-phrase-sense pair. Wordnet has not sufficiently supported English phrases. There are two things that need to be done. First, additional work needs to be done to compose a Wordnet-like thesaurus that support English phrases. Second, the sentence preprocessing process needs to be improved to handle phrases. The process should be able to tokenize a phrase from a document element. Given the two improvements, the issue is expected to be solved.

5 Conclusion

Our work focuses on providing an approach to identify which page number refers to page that is relevant to given indexed term mentioned in a back-of-book index page. The approach relies on measuring the semantic relevancy tween a page (i.e. a sentence in a page that contains an indexed term) and the indexed term using a lexical database. Our experimentation shows that the approach is considerably substantial in term of reliability as a domain expert in page number identification for back-of-book index, with kappa value of 0.6034.

Our future work includes improving the approach to handle phrase in indexed terms and document element. In order to work on this issue, we need to improve a lexical database to handle all phrases and improve the sentence preprocessing process to recognize terms that made of more than one word. Furthermore, future work also includes improving the reliability of the approach by experimenting with different attributes of a page (such as, paragraph, word frequency, weight, etc.) and different approaches (such as, the use of data mining and ontology).

References

1. Diadato, V.: Cross-references in back-of-book indexes. The Indexer 17(3), 178–184 (1991)
2. Lahtinen, T.: Automatic indexing: An approach using an index term corpus and combining linguistic and statistical methods. Report, University of Helsinki, Helsinki (2000)
3. Gottschalg-Duque, C., Lobin, H.: Ontology extraction for index generation. Building Digital Bridges: Linking Cultures, Commerce and Science. In: The 8th ICCC/IFIP International Conference on Electronic Publishing, Brasília - DF, Brazil, pp. 111–120 (2004)
4. Medelyan, O., Witten, I.H.: Thesaurus-Based Index Term Extraction for Agricultural Documents. In: The 6th Agricultural Ontology Service (AOS) Workshop, Vila Real, Portugal, pp. 1122–1129 (2005)
5. Dao, T.N.D., Simpson, T.: Measuring Similarity between sentences, `http://opensvn.csie.org/WordNetDotNet/trunk/Projects/Thanh/Paper/WordNetDotNet_Semantic_Similarity.pdf` (last accessed June 3, 2013)
6. Wu, Z., Palmer, Z.: Verb Semantics And Lexical Selection. In: The 32nd Annual Meeting on Association for Computational Linguistics, pp. 133–138 (1994)
7. Levenshtein, V.: Binary codes capable of correcting deletions, insertions and reversals. Sov. Phys. Dokl. 10, 707–710 (1965)
8. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20(1), 37–46 (1960)
9. Landis, J.R., Koch, G.G.: The Measurement of Observer Agreement for Categorical Data. Biometrics 33(1), 159–174 (1977)

Classification and Clustering Techniques

The Data Analysis of Stock Market Using a Frequency Integrated Spherical Hidden Markov Self Organizing Map

Gen Niina, Tatsuya Chuuto, Hiroshi Dozono, and Kazuhiro Muramatsu

Faculty of Science and Engineering, Saga University,
1-Honjyo Saga 840-8502 Japan
hiro@dna.ec.saga-u.ac.jp

Abstract. In stock market, prediction of the stock price fluctuation has been important for the investors. However, it is hard for the beginner investors to predict the stock price changes due to the difficulty of estimating a company's state makes. To estimate company's state, we propose the suitable model using Spherical-Self Organizing Map that is integrated frequency vector and Hidden Markov Model to estimate hidden state from the time series data. On this paper, the power company stock price movements are used as the time series data, and we also show a result using improved the Self Organizing Map.

Keywords: Self Organizing Map, Hidden Markov model, stock price movements.

1 Introduction

In current capital markets, a variety of time-series information is used to make determinations about the state of investments. However, it is extremely difficult for private investors to analyze such enormous amounts and various types of data to infer the hidden state of a firm. Therefore, in this paper we will propose a learning model which can infer the state of firms and classify them on a map as an approach to more easily analyze information on a firm and determine the state of the firm. As part of this approach, self-organizing maps, which can classify similar types of data on a map using unsupervised learning, and Hidden Markov Self-Organizing Maps, which incorporate Hidden Markov Models to infer hidden states, were used but found to be insufficient for classification [1]. Consequently, we made some improvements to traditional learning models and developed a Frequency Integrated Spherical Hidden Markov Self-Organizing Map (F-S-HMM-SOM), which integrates frequency vectors into the learning nodes of the self-organizing maps in order to take into account the symbol output frequency of time-series data, and we conducted experiments to test whether the classifications – in comparison with traditional learning models – were being done on the basis of a firm's internal state. With the use of F-S-HMM-SOMs, not only can hidden states be extrapolated from given data – self-organization also allows us to obtain intermediate and relative states in the data during the mapping

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 195–204, 2015.
DOI: 10.1007/978-3-662-46742-8_18

process; as a result, it can be expected that we will be able to use a variety of probabilistic models to make inferences.

2 F-S-HMM-SOM

A Frequency-Integrated Hidden Markov Spherical Self-Organizing Map is an improvement on Hidden Markov Self-Organizing Maps, which are spherical self-organizing maps that can visually display data classifications and incorporate Hidden Markov Models (HMMs) into the nodes of the map. In addition to HMMs on the nodes of this map, frequency vectors have been integrated to take into account the symbol output frequency of time-series data. (Fig. 1)

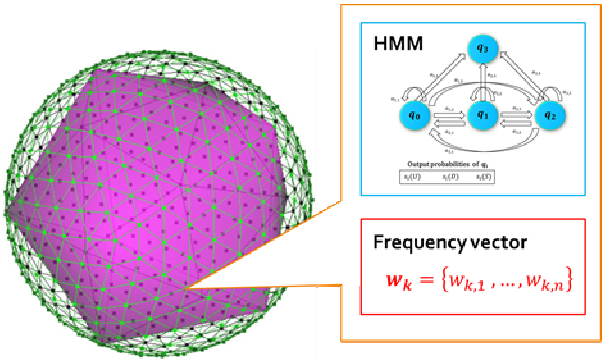


Fig. 1. F-S-HMM-SOM

Furthermore, all nodes on the map have the same structure, the same number of states and symbols contained in each node's HMM, and the same dimensions of frequency vectors for each node. The self-organizing map treats the objects of classification – each of the sets of symbols observed from each of the probabilistic models – as input da-ta. From the input data, each probabilistic model is inferred, and as these models are learned on the map, classification is done on the models inferred from the input data. This learning algorithm is shown below.

- **HMM-S-SOM algorithm**

The symbols used to describe the learning algorithm are defined as follows:

N_k : The k^{th} node on the self-organizing map
θ_k: State transition probability a_{ij} and symbol output probability $S_i(x)$ parameters included in the HMM on the k^{th} node.

Here, a_{ij} refers to the probability of state q_i transitioning to state q_j, and $s_i(x)$ refers to the probability of symbol x being the output from state q_i.

U_y : Refers to the y^{th} set of strings, and is the format of the input data of the F-S-HMM-SOM.

$U_y = \{u_{y,1}, \cdots, u_{y,h}\}$ is the representation where $u_{y,l}$ represents all strings included in the set of strings, and the parameters of the HMM in each node within the neighborhood learning range are updated as the learning takes place to plausibly generate these string sets.

The flow of the learning algorithm is as set out below.

1. Initialization of the Spherical Map

The HMM parameter, Θ_k, and the frequency vectors for node N_k of each sphere are assigned random values on initiation.

2. Calculation

With respect to the input data W_y, the Euclid distance $E_k(y)$ between the likelihood, $L_{20}^k(y)$ of each of the HMM parameters Θ_k at 20 updates using the Baum-Welch algorithm and the frequency vectors $w_k = \{w_{k,1}, \ldots, w_{k,n}\}$ and input frequency vectors $v_y = \{v_{y,1}, \ldots, v_{y,n}\}$ for each node is computed as follows:

$$L_{20}^k(y) = f_j^y(\max t) \tag{1}$$

$$E_k(y) = \sqrt{\sum_i \left(w_{k,i} - v_{y,i}\right)^2} \tag{2}$$

Here, $f_j(t)$ is the Forward algorithm, and can be obtained from the dynamic programming algorithm based on the recursive formula in formula (3).

$$f_j(t) = s_i\left(u_y[t]\right) \sum_{q_i} f_i(t-1) a_{i,j} \tag{3}$$

However, the HMM parameters Θ_k in (3) are virtually replaced by the new parameters, $\overline{a_{i,j}}, \overline{s_i(x)}$ based on the following formulas of the Baum-Welch algorithm [2].

$$\overline{a_{i,j}} = \frac{A_{i,j}}{\sum_{j*} A_{i,j*}} \tag{4}$$

$$\overline{s_i(x)} = \frac{E_i(x)}{\sum_{i*} E_{i*}(x)} \tag{5}$$

The expected values $A_{i,j}$ $E_i(x)$ in (4) and (5) can be obtained from the following formulas:

$$A_{i,j} = \sum_{m=1}^{y} \frac{1}{L_r^k} \sum_t f_i^m(t) \cdot a_{i,j} \cdot s_i(w[t+1]) \cdot b_j^m(t+1) \tag{6}$$

$$E_i(x) = \sum_{m=1}^{y} \frac{1}{L_r^k} \sum_{t:w_m[t]=x} f_i^m(t) \cdot b_i^m(t) \tag{7}$$

Here, $b_j(t)$ is the Backward algorithm and can be obtained from the dynamic programming algorithm based on the recursive formula in formula (8).

$$b_j(t) = \sum_{q_i} b_i(t+1) a_{j,i} \cdot s_j(w[t]) \tag{8}$$

Also for (8), formulas (4) and (5) are followed when there are updates and $b_j(t)$ is calculated using the new parameters.

However, in the program these calculations are done by securing separate memory and substituting virtual parameters so that the calculations do not actually change the parameters in the nodes.

3. Finding the Winner Nodes

However, with respect to the input data, node N_i that fulfills the following formula shall be designated as a Winner node.

$$argmin_i \ \left(1 - L_{20}^i\right) \cdot E_i(y)^* \tag{9}$$

Here $E_i(y)^*$has been normalized so that the maximum value is 1.

4. Updating the Winner Nodes

The HMM parameters of Winner nodes are updated 20 times using formulas (4) and (5) of the Baum-Welch algorithm, and all elements, $w_{k,i}$ of the frequency vector w_k are updated to $\overline{w_{k,i}}$ following the formula:

$$\overline{w_{k,i}} = w_{k,i} + \alpha(v_{y,i} - w_{k,i})$$

Here, α is the learning rate and takes the range $0<\alpha<1$.

5. Updating the Neighbors

Based on the SOM learning algorithm, the parameters of the neighboring nodes approach the parameters of the Winner nodes.

6. Updating the SOM Parameters

As the learning parameters, such as the learning rate and the neighborhood range, are updated in the SOM, steps 2-4 are repeated for the number of times of training. The neighborhood range is designated by its solid angle with the Winner nodes and is made to become smaller with an increase in training (Fig. 2). The learning rate will also become smaller in value as training increases.

- **Consideration of Symbol Output Frequency**

The input frequency vector handled by the F-S-HMM-SOM learning algorithm contains data that has vectorized the symbol output frequency of the symbol sets

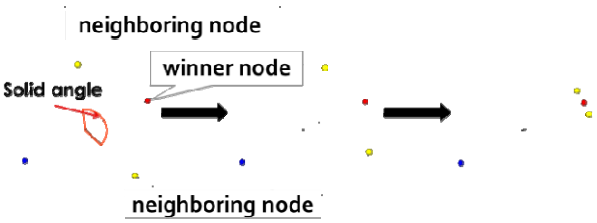


Fig. 2. Neighborhood range

in the input data. The dimensions of the frequency vectors are considered as follows: Assuming three types, A,B, C, of symbol output – for the symbol output by the nth transition of the Hidden Markov Model that generates a symbol for each transition of state, combined with the symbol output by the following transition n+1, there are nine combinations taking into account combinations of two from A,B, C.

Furthermore, for cases in which the transitions are completed after one transition, since only one character is generated, including the three character outputs of A,B,C, there are 12 possible character combinations generated from the pre- and post- state transitions. If the string set of the input vector has symbol length of l, since the output symbols would have a maximum of l state transitions, when the symbols are converted into frequency vectors, $12(l-1)$ becomes the dimensions of the vector. Accordingly, with m types of character, if the maximum length of the symbol set is l_{max}, $m^2 \cdot l_{max} + m$ is the dimensions of the input frequency vector.

For input data with the set of characters {ACCB, BCC}, the conversion from concrete input data to an input frequency vector would be as follows:

① First, a window the length of the two strings is opened and the program checks the window's characters one-by-one, starting from the top of the string, looking for which of the 12 combinations of two characters is in the window. For example, for the string ACCB, the combination of the first two characters is AC, and as shown in Fig. 3, the element AC is set to 1 while the other elements are set to zero, so the string in the first window would be converted to the vector ”000001000000”.

In the same manner, the string CC in the next window would be converted to the vector ”000000000001”, and the string CB in the following window would be converted to the vector ”000000000010. ” In this manner, ACCB is defined as the vector,”000001000000000000000001000000000010.”

Using this method, the string BCC is converted to the vector, ”000000001000000000000001.”

② Taking the sum of all the vectors derived from each of the strings and dividing each element of the summed vector by the number of converted vectors, the frequency vector with respect to the set of strings of the input data is defined by mean of each of the elements (Fig. 4).

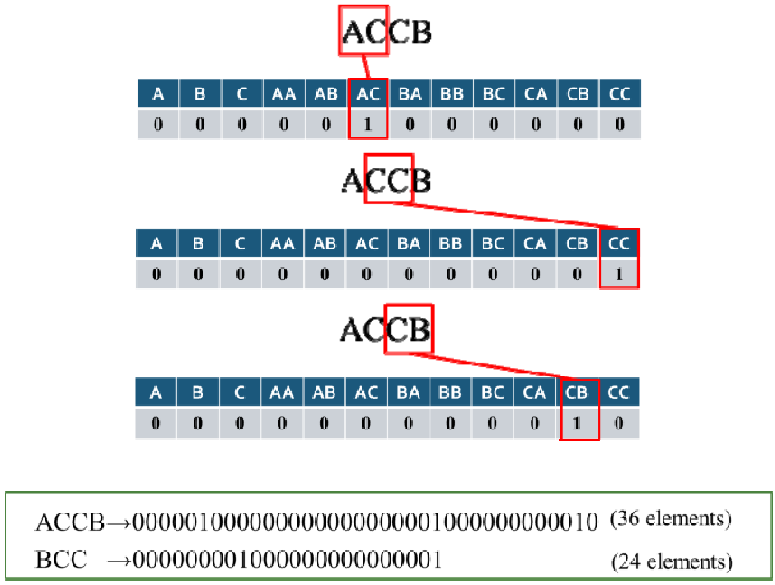


Fig. 3. Conversion to the vector

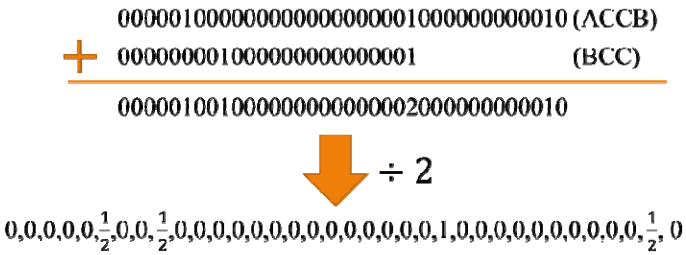


Fig. 4. Conversion of the sequence of symbols to vector

3 Experimental Result by Using F-S-HMM-SOM

In this study, daily stock data of electric power companies (Chugoku and Hokkaido Electric Power Company) was used to classify and infer the state of a firm. We prepared the past 57 days worth of this time-series data, and then prepared 10 days worth of consecutive data over 5 day periods, moving back one day at time, from the first day to the fifth day, followed by the second day to the sixth day, and so on; these 10 days were then set off as one section of data, so that 47 sections of time-series data were extracted. Since these extracted sections were handled as training data, changes in the stock prices for each of the consecutive five day periods in these sections were substituted with symbols —(U) for upward trends, (D) for downward trends, and (S)

for no change — and made into training data. Fig. 5 displays an example of the conversion to symbols for the first section from the first to the tenth day of the data extracted from the time-series data. Additionally, there were four HMM states given as the initial value of the node including the final state, and there were three types of output symbol, U, S, D (Fig. 6).

The display of the spherical map uses an approach known as a U-matrix, which can more visually express differences in parameters between nodes. Using this approach, we expressed differences between parameters in the nodes as mountains and simplified the classification results by changing the color of the mounds on the basis of their height. Stronger shades of red express larger differences in the nodes.

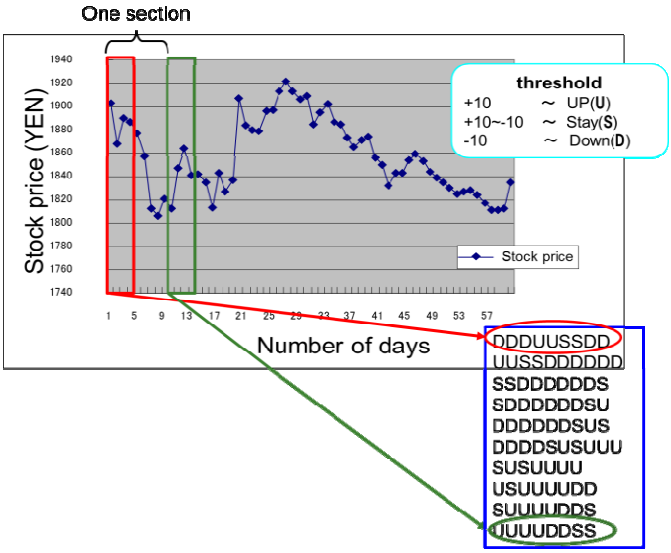


Fig. 5. Conversion example of the stock price to strings

The experimental results in Fig. 7 and Fig. 9 show the classification results of Chugoku Electric Power Company and Hokkaido Electric Power Company, respectively. The red lines on the map express continuity between the time series, and the data is connected in sequential order, from the first day to the last day. Looking at the classification results, it can be seen that the flow of the time-series data has been properly learned and that the lines are smoothly connected without any convoluted intersections. To evaluate whether transitions in the state of a firm were learned, we used the state of the firm — whether good, bad, or stagnant — as our criteria for judgment, and graphed the mean stock price for each section of data to create a visual aid of the stock price fluctuations for each section (Fig. 8, Fig. 10)

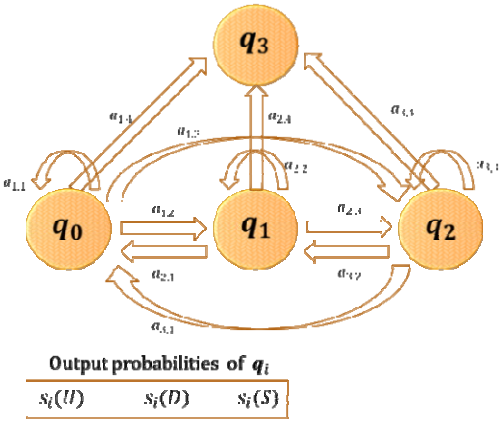


Fig. 6. HMM for price movement of stocks

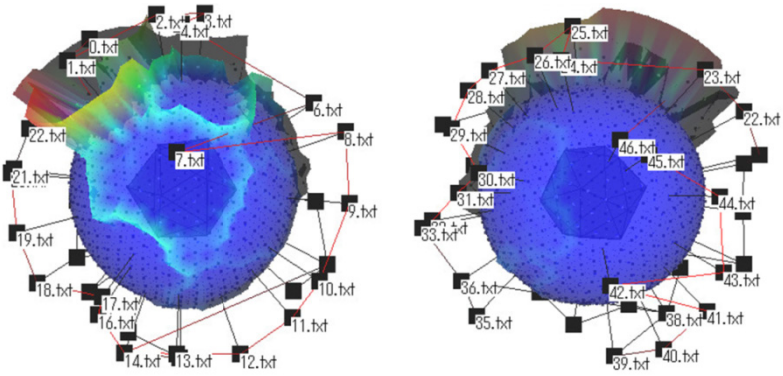


Fig. 7. The mapping results of Chugoku electric power company using F-S-HMM-SOM

With respect to the Chugoku Electric Power Company, the graph in Fig. 8 shows a downward trend in the state of the firm from the first day roughly until the sixth day, and this can be confirmed from the U-matrix on the map.

The mountainous upper region on the map begins to recede from the seventh day onward, properly mapping the gradual change from a downward to an upward trend. From the 25th day onward there is a slight downward shift and the mapped location begins to decline, although we can confirm from the map that it gradually shifts back to an upward trend from around the 37th day.

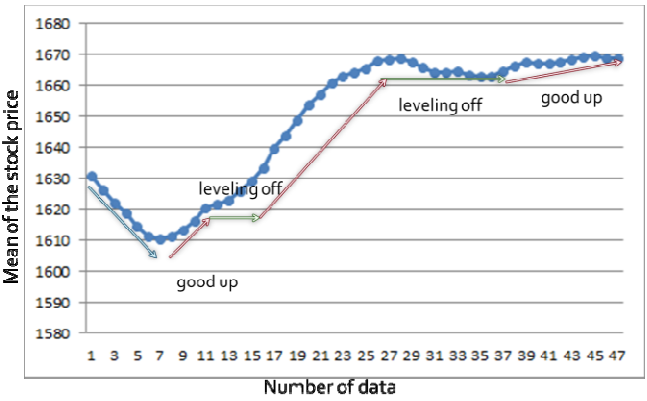


Fig. 8. Mean of the stock price for each data set (Chugoku electric power company)

We were also able to obtain proper learning results with the connection of time-series data with respect to mapping classifications for Hokkaido Electric Power Company. The mapped classifications of the state of the firm show the range of the downward trend lasting from the first day until approximately the 14th day and are expressed in the U-matrix. From the 15th day to the 33rd day, while there were some slight fluctuations in the graph, the overall upward trend of the state of the firm was classified so as to move away from the downward-trending nodes. The point farthest away from the downward trend was mapped on the 33rd day, and from the 33rd day onward the map moves away from the upward trend and results in mapping stagnant-state data to a position that is similar to neither the upward- nor the downward-trending nodes, implying that the state of the firm was classified appropriately.

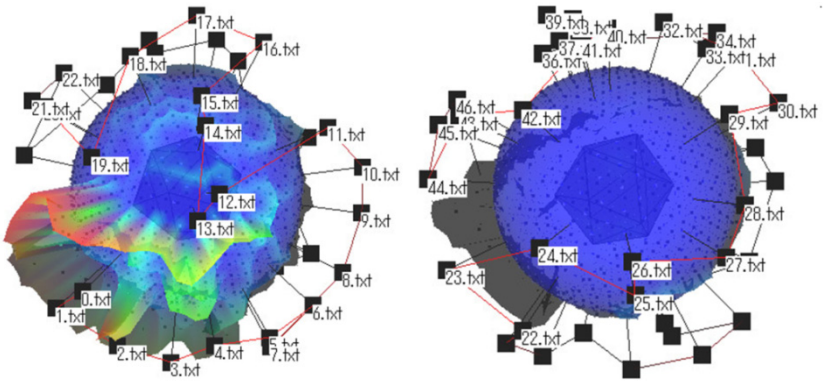


Fig. 9. The mapping results of Hokkaido electric power company using F-S-HMM-SOM

4 Conclusion

Taking into account the frequency of output symbols, the states of firms and connections between time-series were properly classified. Furthermore, the relationship between the distances in the mapped data do not necessarily reflect the degree of similarity between the data, but these relationships at times are mapped so that downward-trending and stagnant trends move away, while at other times they are mapped at the most far-removed point from the downward and upward trends. In the future, it can almost certainly be said that there will be a necessity for the development of a system that automatically color codes these maps; further, there are a variety of points that must be considered to determine whether or not using these kinds of classification results can predict transitions in the state of a firm.

References

1. Niina, G., Dozono, H.: The Spherical Hidden Markov Self Organizing Map for Learning Time Series Data. In: Villa, A.E.P., Duch, W., Érdi, P., Masulli, F., Palm, G. (eds.) ICANN 2012, Part I. LNCS, vol. 7552, pp. 563–570. Springer, Heidelberg (2012)
2. Akuts, T.: Mathematical Models and Algorithms in Bioinformatics. A Series of AlgorithmScience, vol. 12. Kyoritsu Shuppan (2007)
3. Nakatsuka, D., Oyabu, M.: Application of Spherical SOM in Clustering. In: Proc. Workshop on Self-Organizing Maps (WSOM 2003), pp. 203–207 (2003)
4. Yamaguchi, N.: Self Organizing Hidden Markov Models. In: Wong, K.W., Mendis, B.S.U., Bouzerdoum, A. (eds.) ICONIP 2010, Part II. LNCS, vol. 6444, pp. 454–461. Springer, Heidelberg (2010)
5. Jaziri, R., Lebbah, M., Bennani, Y., Chenot, J.-H.: SOS-HMM Self Organizing Structure of Hidden Markov Model. In: Honkela, T. (ed.) ICANN 2011, Part II. LNCS, vol. 6792, pp. 87–94. Springer, Heidelberg (2011)
6. Abe, T., Ikekura, T., et al.: Informatics for unrevealing hidden genome signatures. Genome Res. 13, 693–702
7. Dozono, H., Kabashima, T., et al.: Visualization of the Packet Flows using Self OrganizingMaps. Wseas Transactions on Information Science &Applications 7(1), 132–141 (2010)

Attribute Selection Based on Information Gain for Automatic Grouping Student System

Oktariani Nurul Pratiwi, Budi Rahardjo, and Suhono Harso Supangkat

School of Electrical Engineering and Informatics,
Bandung Institute of Technology,
West Java, Indonesia 40132
oktariani@students.itb.ac.id, br@paume.itb.ac.id, suhono@stei.itb.ac.id

Abstract. Cooperative learning is an approach of learning process together in small groups to solve problems together. Cooperative learning can enhance students' ability higher than individual learning. One of the key that can affect the success of cooperative learning is formation. Heterogeneity in cooperative learning can improve cognitive performance. The problem is hard and need long time to determine students into an appropriate group. A student has many attributes that defined their characteristics from academic factor and non - academic factor, such as motivation in learning, self-interest, learning styles, friends, gender, age, educational background of parents, and other explanation of the uniqueness. The purpose of this study is to determine what is the most influential attribute in grouping process by calculate the information gain of each attributes. Then, we can reduce some attributes. The result of the experiment that the most influential and relevant attribute in the process of formating a group is learning style.

Keywords: Cooperative learning, grouping students, information gain.

1 Introduction

Cooperative learning is an approach of learning process together in small group with members of the group to accept differences and solve problems together [1]. Based on studies, it considered successful learning methods can improve the learning process more effective. The factor that can be affecting the success of cooperative learning process is the composition of students in a group [2]. But, decide formation of a group is not easy. Many teachers divide students into a group randomly. Whereas it can not be done by choose student randomly [1] [3]. It needs heterogeneous characteristics of students. A characteristic of student consist of many attributes like gender, attitude, interest of the material, self-motivation, self-confidence, shyness level and learning performance [4].

Cooperative learning has been extensively researched by many researchers. Rahel Bekele [4] who build systems for supporting cooperative learning by applying automatic clustering process. The process of grouping based on academic factors and non-academic factors. Academic factors consist of the level of school

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 205–211, 2015.
DOI: 10.1007/978-3-662-46742-8_19

facilities, the ability of student to use the language of instruction, the level of motivation in learning, and self-interest on the subject matter. Non-academic factors include gender, age, parents' economic circumstances, educational background of parents, parents' educating style, and an explanation of the uniqueness of the student. Method used in this study is the Bayessian Networks. From these results, the level of accuracy in classifying 79.8% and the ability of students increase 68.29%. Greater than grouping students randomly or classify by themselves.

Furthermore, research study conducted by same researcher [5]. System improved grouping method based on the attributes of student [4]. The system used Ant Colony Optimization method for grouping. It showed the Ant Colony Optimization can give better performance on grouping students than Bayessian Networks.

Shuangyan Liu, et al [2] who build IGLS (Intelligent Grouping based on Learning Styles). IGLS is an online supporting system to automatically grouping students. IGLS classify students based on learning styles. Where students are grouped with other students who have different learning styles. This research has not been tested directly on students.

Many research use different attributes of student to determine the formation of a group. But, if system use too many attributes it can make grouping process longer, even students in large number. It is the reason why we need to decrease number of the attributes by selecting which more relevant.

In this study, we try to find out the most relevant attribute student in learning groups used information gain. We use many attributes of student to decide group formation based on the academic factors, non-academic factors and learning styles of students.

2 Information Gain

Information gain is one of a feature selection method. Feature selection is a heuristic for selecting the splitting criterion that "best" separates a given data partition [6]. This measure is based on work by Claude Shannon on information theory. Information gain is used to induction of decision tree using this formula:

$$Info(D) = -\sum_{i=0}^{m} p_i log2\ (p_i)$$

Info(D) is just the average amount of information needed to identify the class label of tupel in D. The tuples D on some attribute A having v distinct value, {a1, a2, a3, , av}, as observed from the training data. If A is discrete-valued, these values correspond directly to the v outcomes of a test on A. Attribute A can be used to split D into v partitions or subsets, {D1, D2, , Dv}, where Dj contains those tuples in D that have outcome aj of A. Information gain is defined as the difference between the original information requirement and the new requirement.

The attribute with the highest information gain, (Gain(A)), is chosen as the splitting attribute at node N. This is equivalent to saying that we want to partition on the attribute A that would do the "best classification" [6].

3 Student Model

Student model is a model that stores specific information from students to be able to recognize the characteristics of students automatically [11]. In this paper, the student model consist of academic factors, non-academic factors and the learning style of students.

Academic factors consist of the level of motivation in learning, self-interest on the subject matter and information of whose have close relation to the student. Non-academic factors include gender, educational background of parents, and an explanation of the uniqueness of the student. The last is the kind of the learning styles of students.

4 Learning Style

Learning style is kind of style of student when they learn something new. Adjusting learning styles in learning process can make students learn subjects easier [7]. There are many kinds of learning styles. In this study we used learning style as one of the attribute of students. We focused on analyse the VAK learning style. VAK is an acronym for Visual, Aural, and Kinesthetic as describes in Table 1. The investigations about learning styles for a student is important.

Table 1. The VAK Learning Styles [9]

Learning Style	Characteristic
Visual	Preference for using visual resources such as diagrams, pictures and videos.
Auditory	Preference for information that is heard or spoken. Need to talk about situations and ideas with a range of people.
Kinaesthetic	Preference related to the use of experience and practice (simulated or real). It includes demonstrations, simulations, videos and movies of real things, as well as case studies, practice and applications.

Student can know their learning style by answered the VAK questionnaire. The questionnaire has 30 questions that encourage respondents to reflect and answer from within their experience, rather than from hypothetical situations [8]. So, teacher can get students learning styles easily and quickly.

5 Result and Discussion

There are another step to analyse the students' attributes, like pre-test, post-test and collect information of students. The experiment used data from students who answered the VAK questionnaire, fill the form of personal data, pre-test and post-test. The objective of the questions is to get the attributes of students such non-academic factors, academic factors and learning styles of students. Academic factors consist of the level of motivation in learning, self-interest on the subject matter and information of whose have close relation to the student. Non-academic factors include gender, age, educational background of parents, and an explanation of the uniqueness of the student. As well, based on students' VAK learning styles. Learning styles as the way students prefer to process new information including strategies that are consistently adopted to learn [11].

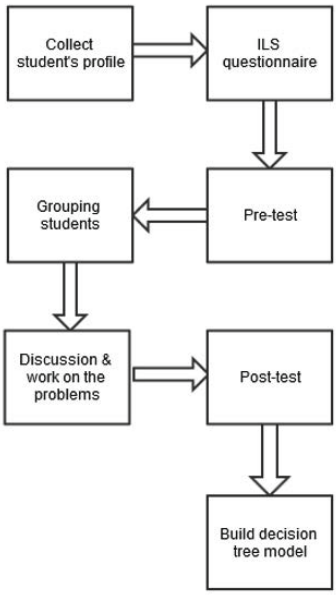


Fig. 1. Research Flowchart

The first phase is to collect data of individual students comprising the students' ability to learn, self-interest on the material, knowing the nearest fellow students, gender, age, educational background of parents, and information about the uniqueness of the student. Then the second stage is to provide VAK questionnaire to determine students' learning styles. The third stage is pre-test used online assessment system to determine the performance of student. After that, teacher classified students into some groups based on their informations.

After each student has a group, they work together to solve problems. Each student can ask, answer and discuss the subject until they can find the answer of

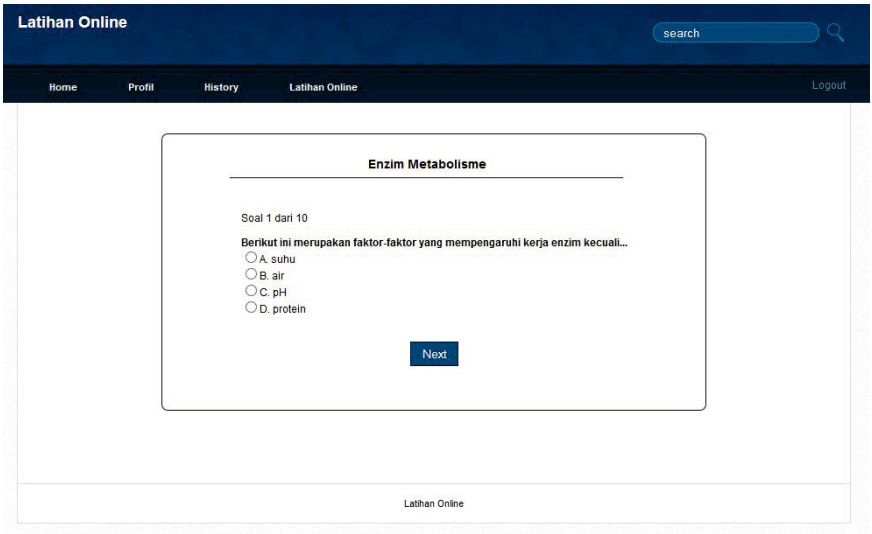


Fig. 2. Online Assessment System in Bahasa

the problem and understand the topic of subject. Teacher can stop them when the time is up to work together.

Students then answer a post-test to measure students' understanding of the material. The results of the post-test is the label that used in ID3 classification models. The questions of post-test are the same as pra-test. Students use an online assessment system (Fig.2) to do both of the test. Labels are used consisted of assessment is very good, good, and less. The third label is derived from the value range of post-test results of students.

The purpose is to determine the attributes that have the highest value of information gain. By calculating the value of the highest information gain, it can be seen that the attributes are not correlated with the success of student learning in groups.

From Table 2, the experiment shows there is an attributes which have the highest value of information gain. That is learning style factor 0.1. Followed by value of information gain of family, motivation and friends factor.

Table 2. Value of Gain

No	Attributes	Gain
1	Live with family	0.06
2	Father educational background	0.044
3	Mother educational background	0.044
4	Interest to the subject	0.024
5	Motivation	0.06
6	Same group as close friends'	0.06
7	Different learning style	0.1
8	Sex	0.025

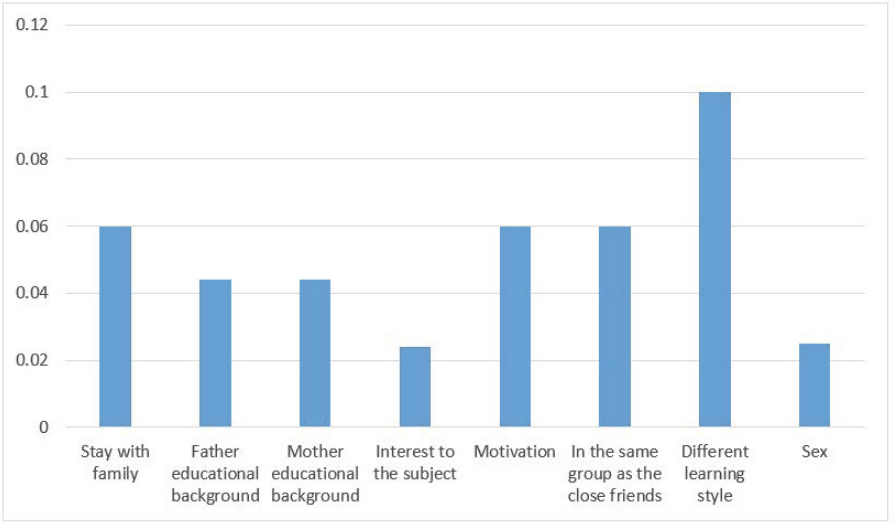


Fig. 3. The most influential attributes in formation of cooperative learning is different learning style

First thing from the experiment, we can see if when students work with their friends who have different learning styles, they can do better too. It showed that if a group formation consists of some students, who have different learning style, it can improve their learning process better.

6 Conclusion and Future Work

In this paper, we present an experiment, that show the most influential and relevant attribute in the process of formating a group is learning style. We can decide the formation of a group based on the learning style first. So, we can implement learning style factor as main factor to decide the composition of formation the group in grouping system.

This experiment still have chance to be expanded. In the future we aim to try implement the model of grouping students to a system. So, the system can grouping automatically to real learners based on the relation attributes that affect the cooperative learning process.

References

1. Dallmann-Jones, A.S.: The Expert Educator: A Reference Manual of Teaching Strategies for Quality Education. Three Blue Herons Pub. (1994)
2. Liu, S., Joy, M., Griffiths, N.: iGLS: Intelligent Grouping for Online Collaborative Learning. In: 2009 Ninth IEEE Int. Conf. Adv. Learn. Technol., pp. 364–368 (July 2009)

3. Felder, R.M., Brent, R.: Cooperative Learning (2007)
4. Bekele, R.: Computer-Assisted Learner Group Formation Based on Personality Traits (2005)
5. Graf, S., Bekele, R.: Forming Heterogeneous Groups for Intelligent Collaborative Learning Systems with Ant Colony Optimization. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 217–226. Springer, Heidelberg (2006)
6. Han, J., Kamber, M.: Data Mining Concepts and Techniques, 2nd edn., pp. 291–310. Morgan Kaufmann Pub. (2006)
7. Graf, S., et al.: In-Depth Analysis of the Felder-Silverman Learning Style Dimensions. Journal of Research on Technology in Education (2007)
8. Fleming, N., Baume, D.: Learning Styles Again: VARKing up the right tree! Educational Developments 7(4), 4–7 (2006)
9. Fleming, N.: How do I learn best?: A students guide to improved learning: VARK-visual, aural, read/write, kinaesthetic. Fleming, Chch, Supplementary website `http://www.vark-learn.com`
10. Kolb, D.A.: Experiental learning: Experience as the source of learning and deelopment. Prentice Hall, New Jersey (1984)
11. Nouh, Y., Karthikeyan, P., Nadarajan, R.: Intelligent Tutoring System-Bayesian Student Model. In: 2006 1st Int. Conf. Digit. Inf. Manag., pp. 257–262 (April 2007)

Data Clustering through Particle Swarm Optimization Driven Self-Organizing Maps

Tad Gonsalves and Yasuaki Nishimoto

Department of Information and Communication Sciences,
Faculty of Science & Technology, Sophia University, Tokyo, Japan
tad-gonsal@sophia.jp, y_nishimoto@sophia.ac.jp

Abstract. The Self-Organizing Map (SOM) is a well-known method for unsupervised learning. It can project high-dimensional data onto a low-dimensional topology map which makes it an efficient tool for visualization and dimensionality reduction. The Particle Swarm Optimization (PSO) is a swarm-intelligence meta-heuristic optimization algorithm based on the feeding behavior of a swarm of birds. In this paper, we have combined these two diverse approaches to form a PSO-SOM which is applied to clustering problems. The advantage of this method is the reduction in computational complexity and increase in clustering accuracy as demonstrated by the experimental results.

Keywords: Self-Organizing Map, Particle Swarm Optimization, machine learning, classification, clustering.

1 Introduction

Artificial neural networks are used for two types of learning – supervised and unsupervised. In supervised training [1, 2], the inputs as well as the outputs are provided. The neural network processes the inputs and compares its resulting outputs against the given outputs. The process of successively adjusting the connection weights (representing the synapsis in the human brain) so as to minimize the error between the given output and the one calculated by the neural net is called supervised learning. The supervised learning algorithm analyzes the training data and produces an inferred function. This function can then be used to classify new datasets. In the unsupervised learning, on the other hand, only the input data are provided without the outputs. Also known as clustering [3], the task of the unsupervised learning algorithm is to group a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in the other groups. Clustering methods include K-means, fuzzy C-means, etc.

The Self-Organizing Map (SOM) introduced by Kohonen [4-6] is a well-known method for unsupervised learning. The SOM map consists of a one or two dimensional grid of nodes, called neurons. It can project high-dimensional data onto a low-dimensional topology map which makes it an efficient tool for visualization and dimensionality reduction. The SOM algorithm is also robust even when the data set is

R. Intan et al. (Eds): ICSIIT 2015, CCIS 516, pp. 212–219, 2015.
DOI: 10.1007/978-3-662-46742-8_20

noisy [7]. Although SOMS can be used in supervised learning [8], they are mostly used in unsupervised learning [9-10]. The SOM methodology has been applied to problems like anomaly detection [11-12] and pattern recognition in diverse fields [13-14].

Particle Swarm Optimization (PSO) is an evolutionary computation method modelled on the intelligent behavior of a swarm [15-17]. In order to find an optimal solution to the problem, PSO uses a number of particles that constitute a swarm flying around in the search space looking for the best solution. Each particle is represented as an N-dimensional point. Each particle renews its flying based on its own past history of flying as well as that of the best particle in the swarm.

The success of the application of PSO for classification using neural nets has been demonstrated in various studies [18-23]. It seems natural to extend the PSO to perform automated clustering tasks due to the manner in which particles of a swarm cluster to similar regions over time. In the literature, there are a number of SOM–PSO applications. In [23], SOM is trained in the usual manner and then the trained weights of SOM are optimized using PSO algorithm. In [24-25], PSO has been used to find the optimum number of clusters of SOM. In [26], PSO is used to predict cluster labels of a trained SOM algorithm.

In this study, we apply the PSO algorithm to the evolutionary learning process of the SOM. Each node in the two-dimensional SOM grid is modelled as a PSO particle which adjusts its weights following the PSO evolutionary mechanism and after several iterations clusters the SOM input data. The motivation for applying the swarm intelligence techniques for the training phase in SOMs is two-fold: (1) to extend the swarm intelligence optimization techniques to diverse fields and (2) to reduce the computational time required for training the dataset.

This paper is organized as follows: Section 2 describes the Self-Organizing Map (SOM). Section 3 describes the Particle Swarm Optimization (PSO) algorithm. Section 4 describes the PSO-SOM novel method for clustering. Section 5 presents the experiments results and Section 6 concludes the paper.

2 Self-Organizing Maps

The Self-Organizing Map (SOM) consists of a one or two dimensional grid of nodes, called neurons. The SOM is an unsupervised learning method used to perform clustering. By using the SOM, one can examine the similarity in the hard to visualize high-dimensional input data. By converting the high-dimensional input data into just one or two dimensional maps, the SOM reveals the trends and correlations in the high-dimensional input data.

The 2-D SOM comprises of a mere two layers - the input layer and the mapping layer (Fig. 1). The input which holds all of the input training data is coupled to all the neurons (nodes) in the mapping layer. Before training the SOM to perform clustering by means of unsupervised learning, the input dataset is divided into training data and test data.

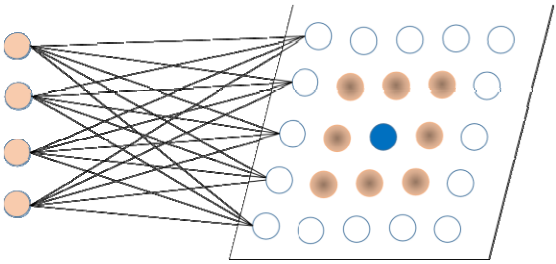


Fig. 1. Self-organizing Map (SOM)

The weights W_{ij} of the neurons in the mapping layer are randomly generated (Equation 1). The input data vectors X_i (Equation 2) from the above training set are fed into the SOM one by one and the distance D between each input data vector and the neurons in the mapping layer is calculated (Equation 3). The neuron closest (with smallest value of D) to the input data vector is called the winner.

$$W_{ij} = (W_{i1}, W_{i2}, \ldots\ldots, W_{in}) \tag{1}$$

$$X_i = (X_1, X_2, \ldots\ldots, X_n) \tag{2}$$

$$D = \| X_i - W_{ij} \| \tag{3}$$

The winner neuron together with the neurons in the immediate neighborhood is updated so as to approach the input data according to Equation (4).

$$W_{ij}(t) = W_{ij}(t\text{-}1) + g * (W_{ij}(t\text{-}1) - X_i) \tag{4}$$

Where, g is a coefficient whose value is given by: $0<g<1$

The training process is iterated for several iterations (epochs). In each iteration, the winning neuron updates itself and its immediate neighbors to approximate the pattern distribution in the input data. After the learning process is complete, the values of the neurons in the mapping layer belonging to the same group come close to one another, i.e., the neurons in the mapping layer arrange themselves into clusters. Thus, the SOM training algorithm obtains a topological ordering of clusters.

3 Particle Swarm Optimization

The Particle Swarm Optimization (PSO) algorithm imitates the information sharing process of a flock of birds searching for food. The population-based PSO conducts a search using a population. Each individual in the population is called the particle and the population as a whole is called the swarm. The fitness function calculated using

the objective function serves as a measure of the performance of the particles. During the search, each particle records two pieces of data: the *pbest* (indicating the particle's local best), and the *gbest* (indicating the global best found by the swarm, so far). In every iteration, the particles renew their search moves taking into consideration their previous *pbest* as well as the swarm *gbest*. The process is repeated for a predefined number of iterations.

The notations used in PSO are as follows: The i^{th} particle of the swarm in iteration t is represented by the d-dimensional vector, $xi(t) = (x_{i1}, x_{i2},\ldots, x_{id})$. Each particle also has a position change known as velocity, which for the ith particle in iteration t is $vi(t) = (v_{i1}, v_{i2},\ldots, v_{id})$. The *pbest* of the i^{th} particle is $pi(t\text{-}1) = (p_{i1}, p_{i2},\ldots, p_{id})$ and the *gbest* (i.e., the particle with the optimal function value found in all the previous iterations), is denoted by the index g. In a given iteration t, the velocity and position of each particle is updated using the following equations:

$$v_i(t) = wv_i(t\text{-}1) + c_1r_1(p_i(t\text{-}1) - x_i(t\text{-}1)) + c_2r_2(p_g(t\text{-}1) - x_i(t\text{-}1)) \quad (5)$$

$$x_i(t) = x_i(t\text{-}1) + v_i(t) \quad (6)$$

where, i =1, 2,…, NP; t = 1, 2,…,T. NP is the size of the swarm, and T is the iteration limit; c_1 and c_2 are constants; r_1 and r_2 are random numbers between 0 and 1; w is the inertia weight that controls the impact of the previous history of the velocities on the current velocity. A large inertia weight facilitates global exploration, while a small one local exploration. Equation 5 is used to compute a particle's new velocity, based on its previous velocity and the *pbest* and the *gbest*. The new velocity is then used to compute the particle's new position (Equation 6).

```
1    Initialize population of particles with random positions & velocities
2    while the termination criterion is not met
3        for each particle i do
4            Compute velocity of the particle using Equation (1)
5            Update position of the particle using Equation (2)
6            Evaluate the fitness
7            Determine pbests & gbest
8        end
9    end
```

Fig. 2. Generic PSO algorithm

4 SOMS Using PSO

The PSO-SOM training algorithm is shown in Fig. 3. The number of neurons (nodes) in the mapping layer is set to m x n. This neuron grid size depends on the size of the input data. The value of the neurons in the mapping layer is chosen randomly from the uniform random distribution (0-1). For each neuron, the input vector is denoted as Lbest. The value of the mapping layer neuron is set as Pbest. The neuron closest to Lbest is called the Firing Particle. We use the Euclidean distance formula to calculate the distance.

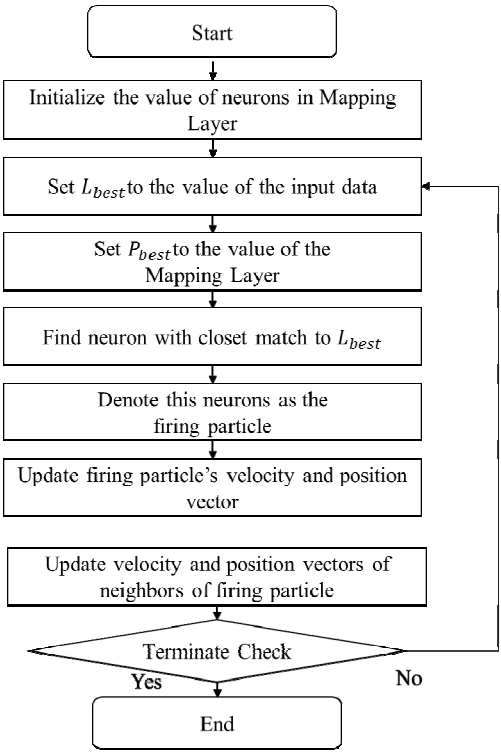


Fig. 3. PSO-SOM training algorithm

$$\text{Firing Particle} = \text{argmin}_i \parallel \text{Lbest-P}_i \parallel \tag{7}$$

This Firing particle's velocity and position are updated along with those of the adjacent neighbors according to Equations 8 and 9.

$$V_i(t+1) = w*V_i(t) + c_1* R_1*(P_{best} - x_i(t)) + c_2* R_2* (L_{best} - x_i(t)) \tag{8}$$

$$x_i(t+1) = x_i(t) + V_i(t+1) \tag{9}$$

where,
i: i^{th} particle
t: generation number
V_i (t): speed of particle i
x_i (t): position of particle i at t generation

P_{best}: optimal solution found by the particle
L_{best}: input value
c_1: self-reliance of the particle
c_2: reliance of the particle on the group best
R_1 , R_2: random number between 0 and 1
w: inertia weight

This training phase is cycled through T iterations. After the end of the training phase, the trained particles are assigned to the respective class. This classification accuracy is then tested on the test data.

5 Experimental Results

We used the PSO-SOM method for clustering the data sets made publicly available in the UCI machine learning repository (http://archive.ics.uci.edu/ml/). The study was conducted on the following sets: Wisconsin, New Thyroid, Pima Indians Diabetes, Wine and Iris. Each data set was divided into a Training set (70%) and a Test set (30%). Adjusting some of the tuning parameters of the PSO algorithm we obtained more than 90% accuracy in the train phase as well as test phase of all the data sets (Table 1).

Table 1. Accuracy of train and test models

Data set	Train Accuracy (%)	Test Accuracy (%)
Wisconsin	93.30	93.42
New Thyroid	93.42	93.67
Pima Indians Diabetes	93.36	93.77
Wine	93.51	93.53
Iris	93.64	93.70

6 Conclusion

In this paper, we have combined these two diverse approaches to form a PSO-SOM which is applied to clustering problems. The Self-Organizing Map (SOM) is a well-known method for unsupervised learning. It can project high-dimensional patterns onto a low-dimensional topology map, making it an efficient tool for dimensionality reduction and visualization. The Particle Swarm Optimization (PSO) is a swarm-intelligence meta-heuristic optimization algorithm based on the feeding behavior of a swarm of birds. The advantage of the hybrid method is the reduction in computational complexity and increase in clustering accuracy as demonstrated by the experimental results.

References

1. Vapnik, V.N.: The Nature of Statistical Learning Theory, 2nd edn. Springer (2000)
2. Gonen, M.: Supervised Multiple Kernel Embedding for Learning Predictive Subspaces. IEEE Transactions on Knowledge and Data Engineering 25(10), 2381–2389 (2013)
3. Kumar, V., Tan, P.: Introduction to data mining, pp. 487–556. Addison Wesley, Pearson International Edition (2006)
4. Kohonen, T.: The SOM Methodology. In: Deboeck, G., Kohonen, T. (eds.) Visual Explorations in Finance with Self-organizing Maps, pp. 159–167. Springer, Berlin (1998)
5. Kohonen, T.: The Self-Organizing Map. Proceedings of the IEEE 78(9), 1464–1480 (1990)
6. Kohonen, T.: Self-Organizing Maps, 3rd edn. Springer, New York (2001)
7. Mangiameli, P., Chen, S.K., West, D.: A Comparison of SOM neural network and hierarchical clustering methods. European Journal of Operational Research 93(2), 402–417 (1996)
8. Ichiki, H., Hagiwara, M., Nakagawa, M.: Kohonen feature maps as a supervised learning machine. IEEE International Conference on Neural Networks 3, 1944–1948 (1993)
9. Chaudhary, V., Ahlawat, A.K., Bhatia, R.S.: An efficient self-organizing map learning algorithm with winning frequency of neurons for clustering application. In: Proc. International Advanced Computing Conference (IACC), pp. 664–668. IEEE (2013)
10. Chaudhary, V., Bhatia, R.S., Ahlawat, A.K.: The selforganizing map learning algorithm with inactive and relative winning frequency of active neurons. HKIE Trans. 21(1), 62–67 (2014)
11. Albayrak, S., Scheel, C., Milosevic, D., Muller, A.: Combining Self-Organizing Map Algorithms for Robust and Scalable Intrusion Detection. In: 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on Computational Intelligence for Modelling, Control and Automation, November 28-30, vol. 2, pp. 123–130 (2005)
12. Sujatha, P.K., Kannan, A., Ragunath, S., Bargavi, K.S., Githanjali, S.: A Behavior Based Approach to Host-Level Intrusion Detection Using Self-Organizing Maps. In: First International Conference on Emerging Trends in Engineering and Technology, ICETET 2008, July 16-18, pp. 1267–1271 (2008)
13. Dragomir, O.E., Dragomir, F., Radulescu, M.: Matlab Application of Kohonen Self-organizing Map to Classify Consumers' Load Profiles. Procedia Computer Science 31, 474–479 (2014)
14. Mohebi, E., Bagirov, A.: A convolutional recursive modified Self Organizing Map for handwritten digits recognition. Neural Networks 60, 104–118 (2014)
15. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, pp. 1942–1948 (1995)
16. Kennedy, J., Eberhart, R.C.: A discrete binary version of the particle swarm algorithm. In: IEEE International Conference on Computational Cybernetics and Simulation, vol. 5, pp. 4104–4108 (1997)
17. Kennedy, J., Eberhart, R.C., Shi, Y.: Swarm Intelligence. Morgan Kaufmann Publishers, San Francisco (2001)
18. Misra, B.B., Dehuri, S., Dash, P.K., Panda, G.: Reduced polynomial neural swarm net for classification task in data mining. In: IEEE World Congress on Evolutionary Computation, pp. 2298–2306 (2008)

19. Khurshid, A.A., Gokhale, A.P.: Classification system for digital signal types using neuro fuzzy system and PSO. In: World Congress on Nature & Biologically Inspired Computing, NaBIC 2009, pp. 373–378 (2009)
20. Dehuri, S., Roy, R., Cho, S.B., Ghosh, A.: An improved swarm optimized functional link artificial neural network (ISO-FLANN) for classification. Journal of Systems and Software 85(6), 1333–1345 (2012)
21. Fernando, L., López, M., Gómez Blas, N., Arteta, A.: The optimal combination: Grammatical swarm, particle swarm optimization and neural networks. Journal of Computational Science 3(1-2), 46–55 (2012)
22. Gonsalves, T., Oguro, A.: Feed-forward Neural Network Training using Particle Swarm Optimization. Advances in Computer Science and Engineering 12(1), 1–13 (2014)
23. Pulido, M., Melin, P., Castillo, O.: Particle swarm optimization of ensemble neural networks with fuzzy aggregation for time series prediction of the Mexican Stock Exchange. Information Sciences 280(1), 188–204 (2014)
24. Xiao, X., Dow, E.R., Eberhart, R.C., Ben Miled, Z., Oppelt, R.J.: Gene clustering using self-organizing maps and particle swarm optimization. In: Proceedings of the second IEEE International Workshop on High Performance Computational Biology, Nice, France (2003)
25. Das, S., Abraham, A.: Automatic kernel clustering with a multi-elitist particle swarm optimization algorithm. Pattern Recognition Letters, 688–699 (2008)
26. Omran, M., Salman, A.: Dynamic clustering using particle swarm optimization with application in unsupervised image classification. Proceedings of World Academy of Science Engineering and Technology, 199–204 (2005)
27. Sharma, A.: Determining cluster boundaries using particle swarm optimization. Proceedings of World Academy of Science Engineering and Technology, 250–254 (2006)

Intelligent Data Processing

A Search Engine Development Utilizing Unsupervised Learning Approach

Mohd Noah Abdul Rahman[1], Afzaal H. Seyal[1],
Mohd Saiful Omar[1],and Siti Aminah Maidin[2]

[1] School of Computing & Informatics, Institut Teknologi Brunei,
Bandar Seri Begawan BE 1410, Brunei Darussalam
{noah.rahman,afzaal.seyal,saiful.omar}@itb.edu.bn
[2] Ministry of Defense, Bolkiah Garrison, Bandar Seri Begawan, Brunei Darussalam
ameena.alhajah@gmail.com

Abstract. This article reports a software development of a generic search engine utilizing an unsupervised learning approach. This learning approach has become apparently important due to the growth rate of data which has increased tremendously and challenge our capacity to write software algorithm and implementation around it. This was advocated as a mean to understand better the flow of algorithm in an uncontrolled environment setting. It uses the Depth-First-Search (DFS) algorithm retrieval strategy to retrieve pages with topical searching. Subsequently, an inverted indexing technique is applied to store mapping from contents to its location in a database. Subsequently, these techniques require proper approach to avoid flooding of irrelevant links which can constitute a poor design and constructed search engine to crash. The main idea of this research is to learn the concept of how to crawl, index, search and rank the output accordingly in an uncontrolled environment. This is a contrast as compared to a supervised learning conditions which could lead to information less overloading.

1 Introduction

The widespread use of internet have revolutionized the way people access to information. The advent of search engines have made these happenings [1,2,3] which have proliferated tremendously over the years. There are approximately 300 million Google users and 2 billion log-in searches per day [4] and few are valuable [5]. These figures are so stupendous which could easily lead to information overloading [6] and vocabulary differences [7].

The core of the searching capabilities is the availability of search engines which represent the user interface needed to permit users to query for information [8]. A basic search engine consists of a crawler, indexer and a searcher or query processor which will finally rank the results according to the required output. A crawler as the name implies, crawl the web site and follows all the links that are available. Developing and running a web crawler is a challenging task as it involves performance, reliability and social issues [9]. It also consumes resources belonging to other organizations as well [10].

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 223–233, 2015.
DOI: 10.1007/978-3-662-46742-8_21

An unsupervised learning approach enables programmers the ability to apply machine learning paradigm to solve uncontrolled variables. This approach is multifaceted and it is incorrect to characterize unsupervised learning as stimulus driven, incremental and passive [11,12,13,14]. This learning method has become apparently important due to the rate of growth of data which has increased tremendously and challenge our capacity to write software algorithm around it. This information overloading has become one of the pressing research problems [15].

In a supervised learning methodology, a model is learned using a set of fully labeled items called the training sets [16]. Past research works have highlighted the importance of comparing multiple learning modes in order to develop more general theories of learning [17,18]. The earliest work by Shepard et al. [19] paved the way for quantitatively modeling of a supervised classification learning.

The unsupervised learning is synonym to Learning with a Teacher [20] or Inductive Machine Learning [21] is to build a generic search engine which can learn the mapping between the input and a non-determined set of outputs which is a total contrast as compared with the supervised learning approach [22]. The unsupervised learning is a multifaceted and performance varies with task conditions [39]. The implementation of this application will enable the next batch of students to clearly understand the working mechanics of a search engine and proceed to the next stage of their learning process.

The organization of this research is sequenced as follow: Firstly, a web page is constructed with an infinite number of links. Secondly, a crawler is developed to exercise the concept of crawling, indexing, searching and ranking. Thirdly, an experiment is conducted to test all searching activities and subsequently leads to discussions and a conclusion is made.

2 Proposed Approach

An uncontrolled environment to develop a search engine is proposed to enable search engineers to understand the algorithm of a generic search engine. A typical search engine is comprised of a world wide web (www), web crawling, indexing and searching [23].

Most of the search engines whether they are used in the industries which still remain as trade secrets for most of their algorithms or academia employ well established algorithms and techniques [24,25]. One of the fundamental problems with scientific uses of commercial search engines is that their results can be unstable [26,27,28,29]. The earliest and the de facto searching approaches provided by the search engine are based upon the Boolean retrieval model [30] so that the search results exceed the users' abilities to absorb [31,32]. Furthermore, there exists a plethora of classification algorithms that are used for supervised learning methodology.

Although most of the current retrieval models are still based upon keywords, many online users are inclined to establish searching process according to the high-level semantic concept [33]. This study utilises the Depth-First Search (DFS) algorithm

retrieval strategy adapted from the graph theory to retrieve pages from sites which are topic-specific or topical searching. This algorithm involves the application of a stack, a data structure concept of last-in, first out (LIFO). This involves the stacking and destacking operations and link lists. The DFS processes each node of binary tree in level by level increasing fashion and left to right in each level [34].

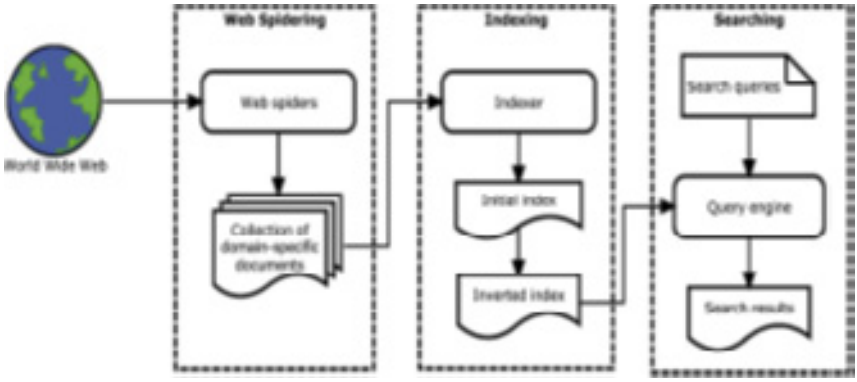


Fig. 1. Overall architecture of a search engine developmental tool [23]

The DFS is established by following hypertext links leading to pages which also allow for easier modification and enhancement of the codes in future. Using this function written in Python, the crawler can keep crawling continuously or deposit or save a list of un-indexed pages for later crawling avoiding the risk of overflowing the stack. This searching mechanism will definitely find a solution if it ever exists [35].

The generic algorithm of a DFS with recursive function is illustrated below:

1. Get the URLs seed.
2. Put them in the stack.
3. For every URL in the stack, download page and extract all URLs in the page.
4. Put those URLs in the stack.
5. Iterate step 3.

This function loops through the list of pages, calling a function *addtoindex* on each one. It then uses Beautiful Soup, which is a Python library available in [36] for pulling data out of HTML and XML files to get all links on that page and adds their URLs to a set called *newpages*. At the end of the loop, *newpages* becomes pages, and the process rolls over again. The function, *isindexed*, will determine if a page has been indexed recently before adding it to *newpages*.

Now that we have learnt how the indexing process works, next is to decide the best algorithm for the indexing process for this project. We will be using inverted indexing which stores a mapping from content such as words or numbers to its locations in a DBMS which offers many perceived benefits [37,40]. Theoretically, creating an inverted index for a set of documents D, each with unique ID *doc_id* involves all codes that are available in [36] written in Python programming language which is concise, easy to extend, multi-platform and it is free.

3 Proposed Approach

Users can used any search engines available to search for any materials or documents. This uniform resource locator (url) is often referred as the anchor tag or the initial url seed. The anchor tag is identified and the subsequent activities will commence until it reaches the final destination of displaying all the links.

In this experiment, we used a PC with a 4th generation Intel Core i5 with a 3.9 GHz processor with 4 GB RAM and 1TB hard drive. Notepad is used as the editor for the coding of HTML, PHP and a Python programming language and MySQL as the database management systems (DBMS) to store the links. For the search engine interface, there should be at least three pages; a crawler's page for crawling URLs and saving them into an index file, a search page for submitting queries and the result page for returning results of the search query. In relation to these, we will also need a website with lots of links to allow the search engine to work and of course, these will require browsers to view the web pages. Any supported browsers will do such as Microsoft IE 7 or above, Mozilla FireFox 3.0 or above, Google Chrome etc.

This section presents the results based on several test cases conducted to monitor the crawler's performance. The first testing is to make sure that the Rainbow website is connected to the internet. It then displays all external links and the results from the crawling process can be generated. These results are analyzed through the use of explanations and screenshots of system behavior and capabilities that forms the basis of the derive core results. The evaluation of the result is done by comparing system output with user expectations using a collection of data and queries.

Fig. 2. shows the buttons which form the basis in understanding the Links, Crawl, Index and Search concepts. The tests results from this figure is shown in the next section.

Fig. 2. Linking, Crawling, Indexing and Searching buttons

3.1 Test Results

● External links - Clicking any button in Fig. 2. navigates users to a webpage that provides links to external websites. Clicking on Email, Dine, Sports and News buttons will show four web pages respectively as shown in Fig. 3. providing contents, images and links to other webpages. This shows that the external links in Fig. 2. are functioning.

Fig. 3. External Links

● Link - The (Link) button in Fig. 2. is used to show all internal links to pages. Unlike the external link, which requires internet connection to get its contents, this link does not require users to be connected to the internet. Clicking the button will bring user to a page providing internal links as in Fig. 4. which again, when clicked goes to another page as in Fig. 5. and goes on until the end of link is reached.

Fig. 4. Internal link level 1

Fig. 5. Internal link level 2

• **Crawled -** The (Crawled) button requires users to view the crawled URLs. When the button is clicked, the page will display URL list showing the total number of crawled URLs coming from the internal as well as the external sites. Testing showed that this function worked successfully because the crawler has managed to list out a total of 19 URLs crawled within 10 seconds as shown in Fig. 6. below.

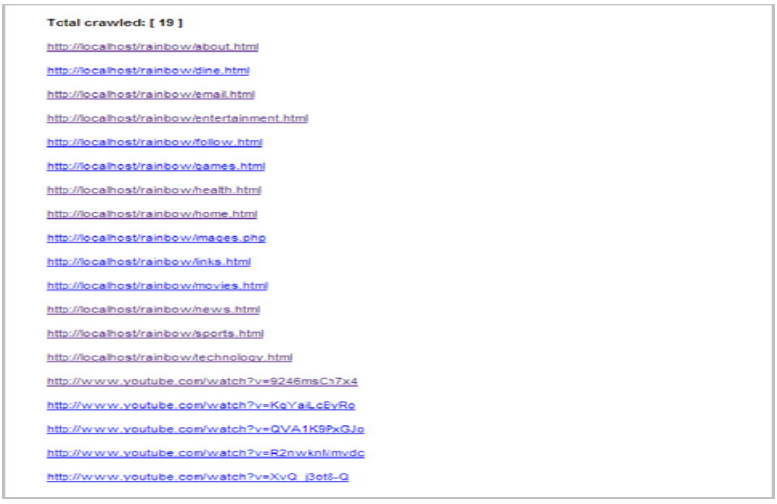


Fig. 6. Crawled URLs lists

● Index - The (Index) button requires users to crawl and index all internal and external websites from Rainbow. When this button is clicked, it will first prompt user

to enter the webpage to be crawled. Having entered the starting page, the crawling and indexing process will start, taking URLs (no words, phrases, images etc.) from every page as shown in Fig. 7., because it was designed to index only links. The indexed URLs is then saved into a 'Pages' table in 'Rainbow' database as shown in Fig. 8., waiting for users to do search. Depending on how big the website is, there could have hundreds of thousands of web pages indexed and then they are all searchable on the website. Testing shows that crawling and indexing process is functioning because the URLs are added to the index table in the MySql database.

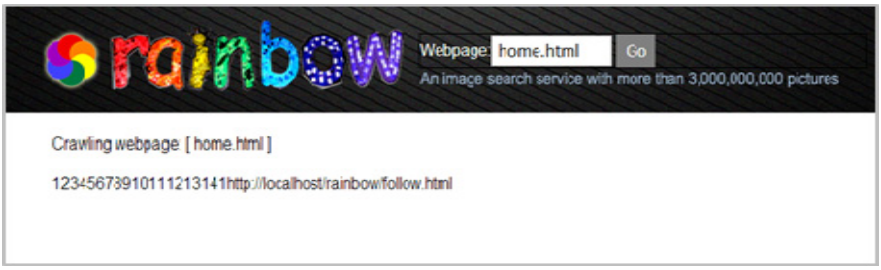


Fig. 7. Crawling and indexing process

● Search - The (Search) button requires users to type in any keyword on the search text box corresponding to the URL name because the indexing process is design to index URLs as keywords. When users perform search, the words they entered in the search box are compared with the indexed URL stored in the database. Results from executing this query will return the following outputs as shown in Fig. 10.

id	title	url	keywords	description
35231		http://localhost/rainbow/about.html		
35232		http://localhost/rainbow/email.html		
35233		http://localhost/rainbow/links.html		
35234		http://localhost/rainbow/images.php		
35235		http://localhost/rainbow/home.html		
35236		http://localhost/rainbow/news.html		
35237		http://localhost/rainbow/technology.html		
35238		http://localhost/rainbow/health.html		
35239		http://localhost/rainbow/entertainment.html		
35240		http://localhost/rainbow/sports.html		
35241		http://localhost/rainbow/dine.html		
35242		http://localhost/rainbow/movies.html		
35243		http://localhost/rainbow/games.html		
35244		http://localhost/rainbow/follow.html		
35245		http://www.facebook.com/ameena.hasasah		
35246		https://twitter.com/Hasanah_Ishak		
35247		http://localhost/rainbow/1.html		
35248		http://localhost/rainbow/crawlList.php		
35249		http://localhost/rainbow/index.php		
35250		http://localhost/rainbow/seaich.php		
35251		http://www.google.com.bn/		
35252		http://www.bing.com/		
35253		http://search.yahoo.com/		
35254		http://www.ask.com/		
35255		http://search.aol.com/aol/webhome		
35256		http://home.mywebsearch.com/		
35257		http://blekko.com/		
35258		http://www.lycos.com/		
35259		http://www.dogpile.com/		
35260		http://www.webcrawler.com/		

Fig. 8. Crawled and indexed URLs saved in database

Based on several outputs from the system, this search engine has proved to be functioning successfully generating outputs matching queries submitted by users with data stored in the index tables. Although this is just a simple search engine with a few basic functions, the crawling, indexing and searching process have produced information to user when searching process is called. In some situations, during searching process, some pages failed to open file contents deriving from the external sites. These problems have affects the effectiveness of the search engine in which the quality of output is not up to the standard of any search engine.

4 Discussion

Based on few test cases, the results show that the quality of output was not up to the standard of effectiveness due to the output quality. This is because the current version of this search engine was designed to index only URLs (links or pages) no words, titles, images, videos or anything else to be used as keywords for the search process. Using URL's resource name alone is not good enough as they are poor descriptions of actual information needs. This could be improved and there are many ways for improvement to produce better results and one way is to have the document indexed on its contents, titles, metadata, images, links etc. so that user can have a broad view of the keywords to be searched. Since this current version of search engine is only used for learning process, in future it can be utilized further to enable document to be indexed on title, images or words, links etc. on the page.

Processing time is the time taken to crawl and index websites whereas response time is a delay between a user submitting a query and the time when results are received, usually measured in milliseconds (ms). Processing and response time are again determined by few things such as the speed of the processor and how huge the index file is. For this research project, we have tested the search engine using a 3.96 GHz Intel processor with 4 GB RAM and it came out that the processing time for the crawling and indexing processes are good enough as compared to our previous testing using the supervised learning approach [18]. This is because the indexing algorithms for this crawler's version are made simple by indexing on multiple fields. If the indexing is designed to do a full-text search or other types of indexing such as titles, images, videos, links etc. this would take longer time which again depends on the size of the site contents.

This search engine is able to crawl the web pages for 35 minutes and downloaded 29853 links from the Rainbow website. In order to maintain a search engine corpus of say, ten billion web pages, in a reasonable state of freshness, say with pages being refreshed every 4 weeks on average, the crawler must download over 4,000 pages per second [38]. This can be achieved using distributed crawlers over multiple computers and each crawling machine must pursue multiple downloads in parallel and this would likelihood overload and crash that web server. Therefore, the politeness policies are suggested to be implemented to limit the amount of traffic directed to the web server.

The performance of the Rainbow search engine is far slower because it was not designed for a distributed crawling. If it is to crawl continuously until the end, it would probably take hours or even days to finish crawling because it involves crawling to external sites as well. Performance will only be slow in areas where indexing images, videos or full-text are involved but this crawler only use URLs as the index field. Therefore, this search engine should perform faster than expected. On the other hand, the response time taken when submitting queries and receiving results is quite fast because the table has been indexed which makes it easy to retrieve specified URLs.

The strength of this extended research of a search engine can be seen in its capability to accomplish the crawling, indexing and searching process for information. Based on our previous discussions, there are rooms for improvement in order to make the search engine more powerful and functional by first implementing the ranking algorithm which was not included in the system. One of the most important steps in improving the website ranking in any search engine is to ensure that it contains plenty of rich information that includes relevant keywords that indicate the subject matter of the page's content. Subsequently, we can suggest to improve the performance of the search engine by implementing certain strategies.

For system effectiveness, that is improving the quality of output, we used headings to summarize page contents and include accurate description of page content in metadata. Apart from these, full-text search can also be used that is breaking a block of unstructured text into individual words, phrases, and special tokens. With these, a wide category of indexed keywords are stored in the index file to enable users to find relevant searches which are far better than using only descriptive URLs as keyword as in this crawler. Besides splitting documents into keywords, ranking can also improve the search quality. This can be done in many ways and the most common ones are using the location and frequency of keywords on the web page. Those with higher frequency are deemed to be more relevant than those with lower frequency.

5 Conclusion

This study has demonstrated a generic search engine which forms the core of applications that facilitate information retrieval process in an unsupervised learning environment. It employs a DFS algorithm for retrieval strategy which allows future code modifications and enhancements. This proposed approach uses all available computing resources efficiently as compared to most tasks performed by high end servers.

The development for this prototype of a search engine will enable students to develop and enhance their programming language and algorithms analytical skills to the next level of their learning process. This will further enhance their abilities to produce new algorithms and laid a ground-work in innovation and creativity.

One of the recommendations can now be made about this search engine is to update regularly the algorithm and need to be checked and revised periodically. Future direction of this study can be further extended using other learning paradigm such as semi-supervised learning or other hybrid learning approaches. A majority of human learning activity can be classified as unsupervised.

References

1. Yahoo! Inc., http://www.yahoo.com
2. Bing Inc., http://www.bing.com
3. Google Inc., http://www.google.com
4. Mishra, A.A., Kamat, C.: Migration of Search Engine Process into the Cloud. International Journal of Computer Application 19(1) (April 2011)
5. Glover, E.J., Lawrence, S., Gordon, M.D., Birmingham, W.P., Giiles, C.L.: Web Search Your Way. Communications of the ACM 44(12), 97–102 (2001)
6. Chen, H., Buntin, P., Sutjahjo, S., Sommer, C., Neely, D.: Expert, prediction, symbolic learning and neural networks: An experiment on Greyhound racing. IEEE Expert. 9(21), 21–27 (1994)
7. Fumas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The Vocabulary Problem in Human-System Communication. Communications of the ACM 30(11), 964–971 (1987)
8. Kumar, G., Duhan, N., Sharma, A.K.: Page Ranking Based on Number of Visits of Links of Web Pages. In: Computer & Communication Technology (ICCT), pp. 11–14 (2011)
9. Brin, S., Page, L.: The anatomy of a Large-Scale Hypertextual Web Search Engine. In: World Wide Web Conference (WWW 1998), pp. 107–117 (1998)
10. Koster, M.: Robots in the web: threat or treat. ConneXions 4(4) (April 1995)
11. Berry, D.C., Dienes, Z.: Implicit learning: Theoretical and empirical issues. Erlbaum, Hillsdale (1993)
12. Hayes, N., Broadbent, D.E.: Two modes of learning for interactive tasks. Cognition 24, 249–276 (1988)
13. Hock, H.S., Malcus, L., Hasher, L.: Frequency discrimination: Assessing global and elemental letter units in memory. Journal of Experimental Psychology, Learning, Memory & Cognition 12, 232–240 (1986)
14. Kellog, R.T.: When can we introspect accurately about mental processes. Memory & Cognition 10, 141–144 (1982)
15. Bowman, C.M., Danzig, P.B., Manber, U., Schwartz, F.: Scalable Internet Resource Discovery: Research Problems and Approaches. Communications of the ACM 37(8), 98–107 (1994)
16. Croft, W.B., Metzler, D., Strohman, T.: Search Engines: Information Retrieval in Practice, p. 344. Pearson Education Inc. (2010)
17. Love, B.C., Markman, A.B., Yamauchi, T.: Modeling classification and inference learning. In: Fifteenth National Conference on Artificial Intelligence, pp. 136–141. MIT Press, MA (2000)
18. Yamauchi, T., Love, B.C., Markman, A.B.: Learning non-linearly separable categories by inference and classification. Journal of Experimental Psychology: Leraning, Memory & Cognition 28, 585–593 (2002)
19. Sheperd, R.N., Hoyland, C.L., Jenkims, J.M.: Learning and memorization of classifications. Psychological Monographs 75(13, Whole No. 517)
20. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall PTR (1998)
21. Kotsiantis, S.: Supervised Machine Learning: A Review of Classification Techniques. Informatica Journal 31, 249–268 (2007)
22. Rahman, M.N., Seyal, A.H., Maidin, S.A.: Search engine development: Adaptation from supervised learning methodology. In: Fourth International Conference on Digital Information Processing & Communications, March 18 -20, pp. 35–42 (2014)

23. Chau, M., Wong, C.H.: Designing the user interface and functions of a search engine development tool. Decision Support Systems 48, 369–382 (2010)
24. Salton, G.: Automatic Text Processing. Addison-Wesley, Reading (1989)
25. Faloutsos, C.: Access Methods for Text. ACM Computing Surveys 17(1), 48–74 (1985)
26. Bar-Ilan, J.: Search engine results over time: A case study on search engine stability, http://www.cindoc.csis.es/cybermetrics/articles/v2i1p1.html (retrieved January 26, 2014)
27. Bar-Ilan, J.: Search engine ability to cope with the changing web. In: Levene, M., Poulovasilis, A. (eds.) Web Dynamics. Springer, Berlin (2004)
28. Mettrop, W., Nieuwenhuysen, P.: Internet search engines - fluctuations in document accessibility. Journal of Documentation 57(5), 623–651 (2001)
29. Rousseau, R.: Daily time series of common single word searches in AltaVista and NorthernLight. Cybermetrics 2(3) (1999)
30. Frants, V.I., Shapiro, J., Taksa, I., Voiskunskii, V.G.: Boolean Search: Current State and Perspectives. Journal of the American Society of Information Science 50(1), 86–95 (1999)
31. Jung, S., Herlocker, J.L., Webster, J.: Click Data as Implicit Relevance Feedback in Web Search. Information Processing & Management 33, 791–807 (2007)
32. Chau, M., Chen, H., Qin, J., Zhou, J.Y., Qin, Y., Sung, W.K., McDonald, D.: Comparison of two approaches to building a vertical search tool: a case study in the nanotechnology domain. In: 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 135–144. ACM (July 2002)
33. Yi, J.: The Research of Search Engine Based on Semantic Web. In: International Symposium on Intelligent Information Technology Workshop (IITAW), pp. 360–363 (2008)
34. Brassard, G., Bratley, P.: Fundamentals of Algorithms, 1st edn., pp. 303–305. PHI Publications, New Delhi (2008)
35. Russel, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River (1995)
36. Segaran, T.: Programming Collective Intelligence: Building Smart Web 2.0 Applications, 1st edn. O'Reilly Media Inc. (2007)
37. Rahman, M.N., Seyal, A.H., Mohamed, H.Y., Mashud, I.: A theoretical framework on the use of database management systems. Journal of Technology & Management 5(1), 36–48 (2007)
38. Najork, M.: Web Crawler Architecture. Encyclopedia of Database Systems (2009)
39. Love, B.C.: Comparing supervised and unsupervised category learning. Psychonomic Bulletin & Review 9(4), 829–835 (2002)
40. Rahman, M.N., Seyal, A.H., Mohamed, H.A.Y.: An empirical framework of DBMS usage in Brunei Darussalam. In: The Fifth Annual Global Information Technology Management World Conference (GITM), San Diego, California, USA, June 13-15, pp. 189–192 (2004)

Handling Uncertainty in Ontology Construction Based on Bayesian Approaches: A Comparative Study

Foni Agus Setiawan[1], Wahyu Catur Wibowo[1], and Novita Br Ginting[2]

[1] Department of Computer Science, Universitas Indonesia, Depok, Indonesia
foni.agus@ui.ac.id, wibowo@cs.ui.ac.id
[2] Department of Informatics Engineering, Universitas Ibn Khaldun Bogor, Bogor, Indonesia
novitaviedie@yahoo.com

Abstract. Ontology is widely used to represent knowledge in many software applications. By default, ontology languages such as OWL and RDF is built on discrete logic, so that it can not handle uncertain information about a domain. Various approaches have been made to represent uncertainty in ontology, one of which with a Bayesian approach. Currently, there are four published approaches: BayesOWL or OntoBayes, Multi-Entity Bayesian Networks (MEBN), Probabilistic OWL (PR-OWL), and Dempster-Shafer Theory. This paper provides a comparative study on those approaches based on complexity, accuracy, ease of implementation, reasoning, and tools support. The study concluded that BayesOWL is the most recommended approach to handle uncertainty in ontology construction among others.

Keywords: Bayesian Network, Ontology, Semantic Web, Uncertainty.

1 Introduction

Ontology is a body of knowledge to explain about a domain. Ontologies or semantic networks are the foundation of the semantic web that widely used in knowledge representation. They facilitate the definition and conceptual characterization of classes, individuals, and their relationships within a domain [1]. Ontology languages in the semantic web, such as OWL and RDF(S), are based on crisp logic and thus can not handle incomplete or partial knowledge about an application domain. However, uncertainty exists in almost every aspects of ontology engineering [2].

Uncertainty often involves the existence of multiple conflicting possible instantiations (i.e. states) of a domain. Where uncertainty exists, we would like to know the likelihoods of the possible instantiations, so that we can reduce or prioritize the set of instantiations to be considered [1]. For example, in ontology reasoning, one may want to know not only if A is related to B, but also how close of A is to B; or, one may want to know the degree of similarity even if A and B are not related to each other. Uncertainty becomes more prevalent in concept mapping between two ontologies, generally happened in web environment, where it is often the case that a concept defined in one ontology can only find partial matches to one or more concepts in another ontology [3].

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 234–246, 2015.
DOI: 10.1007/978-3-662-46742-8_22

Literature study has been carried out by searching for relevant researches about uncertainty in ontology. These approaches can be grouped into two: 1) Using Bayesian probability or networks; and 2) Using fuzzy logic.

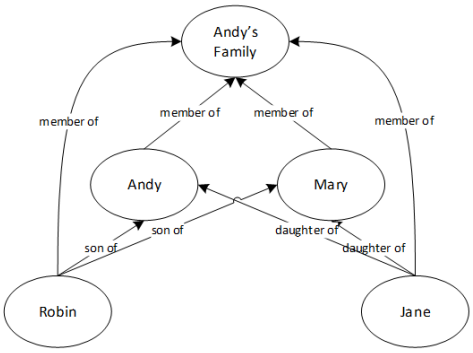
Bayesian Networks (BN) are one of Directed Acyclic Graphs (DAG) model that use a set of random variables and probabilities function applied in DAG to model the relationship between nodes, whereas fuzzy formalisms allow the representation and the gradual assessment of truth about vague information [4]. BN uses graphs to model the knowledge, while fuzzy uses charts to achieve this functionality. This study then focused to observe approaches based on Bayesian only.

Models of Bayesian-based solutions for handling uncertainty in ontology use variety of approaches. These approaches need to be compared and assessed to see the advantages and disadvantages of each, the possibility of technical implementation, as well as compliance with a case that can be handled. Thus, the recommendation can be used as a reference in selecting the appropriate model to solve a case.

2 Literature Review

2.1 Ontology

Ontology is an explicit specification of a conceptualization or formal representation of knowledge. Ontologies provide a definition (specification) of concepts, relationships, and other features related to modeling a realm (domain). It consists of resources with their characteristics and relationships between them. With these specifications, the ontology can be used to check the consistency and reasoning (reasoning) to obtain new facts (formerly implicit becomes explicit) on a sphere. An ontology is not only a classification, or taxonomy of general concepts, it is also a model that includes and maintains the most common properties of concepts, their relationships, existing alignments and known semantics [5].



(a)

```
<People rdf:ID="Andy">
    <memberOf>AF</memberOf>
</People>
<People rdf:ID="Mary">
    <memberOf>AF</memberOf>
</People>
<People rdf:ID="Robin">
    <memberOf>AF</memberOf>
    <sonOf>Andy</sonOf>
    <sonOf>Mary</sonOf>
</People>
<People rdf:ID="Jane">
    <memberOf>AF</memberOf>
    <daughterOf>Andy</daughterOf>
    <daughterOf>Mary</daughterOf>
</People>
<Family rdf:ID="AF">
    <hasName>Andy's Family</hasName>
    <hasMember>Andy</hasMember>
    <hasMember>Mary</hasMember>
    <hasMember>Robin</hasMember>
    <hasMember>Jane</hasMember>
</Family>
```

(b)

Fig. 1. Representation of an ontology by using semantic web (a) and RDF notation (b)

Representation of an ontology is usually depicted in two forms. First, to describe the structure in a simple form is by using graphs of semantic web. Second, in the form of knowledge base used by the software represented using OWL and RDF notation. The examples below illustrate the two forms of representation of an ontology.

2.2 Bayesian Network

Bayesian Network (BN) is a network consisting of nodes and relationships between nodes (arcs) depicted in the form of Directed Acyclic Graph (DAG). Nodes represent facts, classes, or objects, while directed arcs describe unidirectional relationship between the parent nodes and child nodes. The relationship can describe how much the parent influences to the child (influential relations) or how strong the parent incident causes the child (causal relations). The relationship is expressed in weight or degree in a range of numbers 0..1.

Degree or the connection weights are presented in a Conditional Probability Table (CPT) P (Xi | πi), with Xi is any node in the network and πi is the set of parents of Xi. The joint probability distribution of X = (X1,…, Xn) can be calculated as the product of CPT using rules:

$$P(X_i, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\pi_i) \quad (1)$$

BN is widely used in applications to provide a solution regarding to uncertainty or probability in the knowledge domain, such in expert systems, decision support systems, and modeling [13]. Figure 2 below is an example of BN to model relationship of Sparks and Battery to Electrical problem.

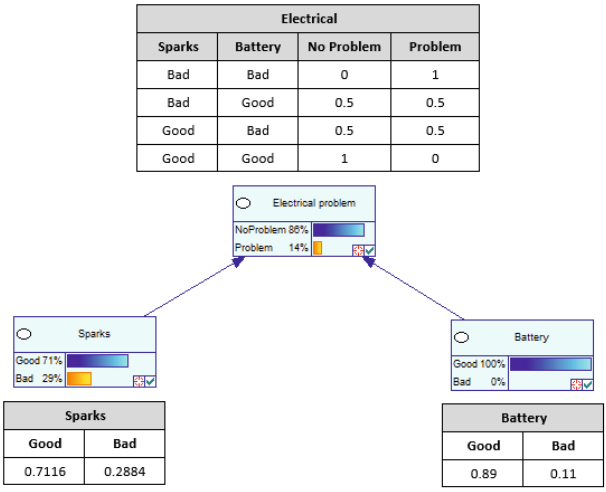


Fig. 2. A Bayesian network example with its CPT

2.3 BayesOWL

BayesOWL is a framework that was first introduced by Ding et. al. (2004) [3]. The framework architecture contains three main components: OWL Parser, BN Structure Constructor, and CPT Constructor.

At first, ontology and uncertainty knowledge entered into the system. OWL Parser consists of two sub-components, namely: Taxonomy Parser (T-Parser) and Probability Parser (P-Parser). T-Parser will translates ontology into taxonomy, then BN Structure Constructor will construct DAG of BN from this taxonomy. P-Parser process the uncertainty of knowledge and collaborates with CPT Constructor construct CPT for the translated BN. Control nodes are created during the construction of BN to facilitate modeling relations among class nodes that are specified by OWL logical operators, and there is a converging connection from each concept nodes involved in this logical relation to its specific control node. There are five types of control nodes in total [2], which correspond to the five types of logical relations:

AND	(owl:intersectionOf)
OR	(owl:unionOf)
NOT	(owl:complementOf)
DISJOINT	(owl:disjointWith)
SAME AS	(owl:equivalentClass)

The nodes in the DAG obtained from the structural translation step can be divided into two disjoint groups: X_R, nodes representing concepts in ontology, and X_C, control nodes for bridging logical relations.

Figure 3 below gives an example diagram of uncertainty handling using BayesOWL framework [2]. In this diagram, X_R={Animal, Male, Female, Human, Man, Woman} while X_C={Bridge_Union, Bridge_Disjoint, Bridge_Intersection_1, Bridge_Intersection_2}. Nodes in X_C are addition nodes provided by the framework to handle uncertainty in the domain.

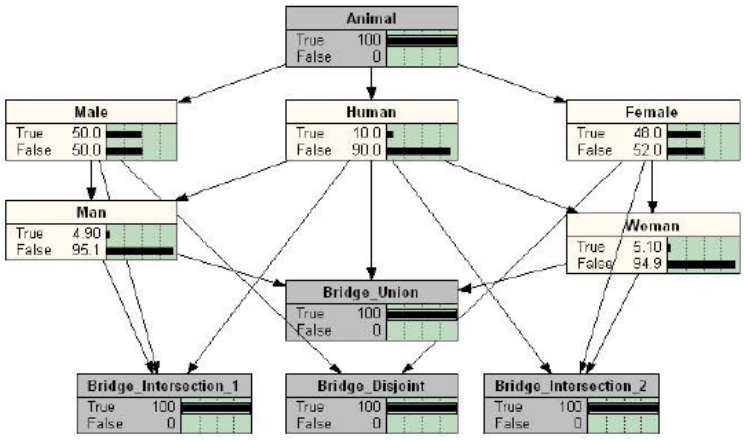


Fig. 3. An example of ontology represented using BayesOWL framework (*Source:* Ding et al., 2005)

The CPT for a control node in X_C can be determined by the logical relation it represents. When all the control nodes' states are set to "True", the corresponding logical relation holds among its parent nodes and all the logical relations defined in the original ontology are held in the translated BN. An algorithm named D-IPFP [3] was
developed to approximate CPTs for each node in XR based on the "iterative proportional fitting procedure" (IPFP), a well-known mathematical procedure that modifies a given distribution to meet a set of probabilistic constraints while minimizing I-divergence to the original distribution [2].

2.4 Multi-Entity Bayesian Networks (MEBN)

MEBN was first introduced by Laskey (2008) in [7]. MEBN is a knowledge representation formalism that combines the power of first-order logic with uncertainty. MEBN provides syntax, a set of model construction and inference processes, and semantics that together provide a means of defining probability distributions over unbounded and possibly infinite numbers of interrelated hypotheses [8]. MEBN theories use directed graphs to specify joint probability distributions for a collection of related random variables. Like logic-based languages, MEBN can represent uncertainty about the values of n-ary functions and relations in a natural way.

The fundamental unit of representation in MEBN is the MEBN Fragment (MFrag). An MFrag is a fragment of a graphical model that is a template for probabilistic relationships among instances of its random variables [9]. Random variables in an MFrag can contain ordinary variables which can be instantiated for different domain entities. An MFrag can be instantiated multiple times by binding its arguments to different entities. MEBN thus provides a compact language for expressing complex graphical models with repeated structure [7].

For example, the MFrag depicted below was taken from a MEBN model on the Star Trek domain, and represents knowledge about the degree of danger to which one's starship is exposed [10]. The fragment graph has seven nodes. The four nodes at the top of the figure are context nodes; the two shaded rectangular nodes below the context nodes are the input nodes; and the bottom node is a resident node.

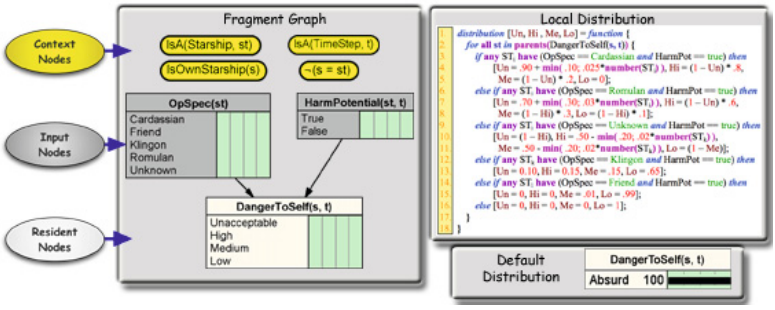


Fig. 4. An example of ontology represented using MEBN framework (*Source: http://www.pr-owl.org/mebn/mfrags.php*)

2.5 Probabilistic OWL (PR-OWL)

PR-OWL is a probabilistic approaches that rely on first-order logic. PR-OWL is an upper ontology [22] enhancement from MEBN that was first introduced by da Costa et. al. (2008) in [11]. PR-OWL was developed as an extension enabling OWL ontologies to represent complex Bayesian probabilistic models based on different probabilistic technologies (e.g. PRMs, BNs, etc.) [12]. PR-OWL designed flexible enough to be used by diverse Bayesian probabilistic tools (e.g. Netica, Hugin, Quiddity*Suite, JavaBayes, etc.). Furthermore, PR-OWL proclaimed to be a framework for developing even the complex probabilistic models.

Reification of relationships in PR-OWL is a necessary thing because of the fact that properties in OWL are binary relations (i.e. link two individuals or an individual and a value), while many of the relations in a probabilistic model include more than one individual (i.e. N-ary relations) [11]. These differences lead to differences in a way of representing relationships, its handling, and the process of reasoning. The use of reification for representing N-ary relations on the Semantic Web is covered by a working draft from the W3C's Semantic Web Best Practices Working Group [13].

Figure 5 below describes an overview of a PR-OWL MEBN theory concepts.

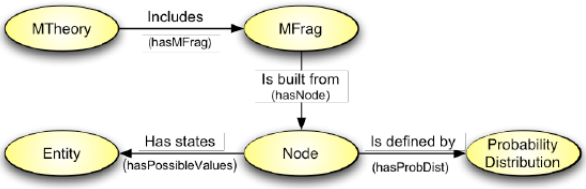


Fig. 5. An overview of a PR-OWL MEBN theory concepts (*Source*: Cesar et al., 2008)

PR-OWL has improved from version 1 in 2008 [11] to version 2 in 2009 [14]. Enhancement has been made so that the PR-OWL 2 can be compatible with OWL as the basic language and has been standardized. Another major difference between PR-OWL 1 and PROWL 2 is that the separate definition of entity in PR-OWL is replaced by OWL's built-in notion of classes and data types [15]. That is, a PR-OWL entity is now identified with either a class or a data type in OWL.

2.6 Dempster-Shafer Theory

Dempster-Shafer Theory [16], also known as theory of evidence or theory of belief function was developed by Arthur P. Dempster and Glenn Shafer in 1975. Dempster-Shafer theory is a mathematical theory of evidence based on belief functions and plausible reasoning (function beliefs and thoughts that make sense), which is used to combine separate pieces of information (evidence) to calculate the likelihood of an event. The idea is that probabilistic approaches reason only on singletons. On the contrary, Dempster-Shafer theory enables us not only to affect belief on elementary hypotheses but also on composite ones.

In probability theory, there is a strong relation between an event and its negation, since its sum equals to 1. But it is not the case in the evidence theory. The evidence theory implies no relations between the existences or not of an event. Thus it models only the belief associated to a class, without influencing the belief allots to others classes [16].

In general, Dempster-Shafer propositions written in an interval: `[Belief, Plausibility]`, where Belief ≤ Plausibility. Belief (Bel) is a measure of the strength of evidence (symptoms) in support of a subset. If the value is 0 then it indicates that there is no evidence, and if the value of 1 indicates the presence of certainty. Plausibility (Pl) is denoted as: `Pl(s) = 1 - Bel(¬s)`. Plausibility also worth 0 to 1. If we believe will ¬s, it can be said that Bel (¬s) = 1, and Pl (¬s) = 0. Plausability will reduce the confidence level of evidence.

Dempster-Shafer theory allows combining evidence from different sources and arriving at a degree of belief (represented by a belief function) that takes into account all the available evidence. In this formalism, a degree of belief (also referred to as a mass) is represented as a belief function rather than a Bayesian probability distribution. Probability values are assigned to sets of possibilities rather than single events. Its naturally encode evidence in favor of propositions.

The following figure taken from [4], illustrates the generic global structure of this ontology called DS-Ontology, in regards of the different classes, hierarchical relations (has subclass) and object/datatype properties (dotted arrows).

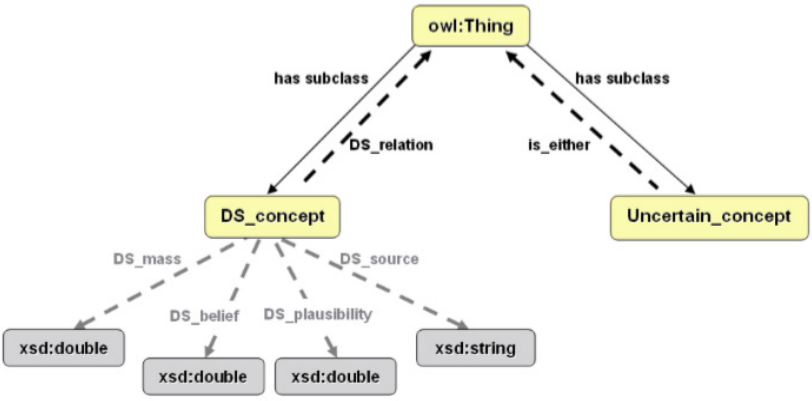


Fig. 6. DS-Ontology structure (*Source*: Bellenger, 2010)

Regarding ontology, Dempster-Shafer theory has been most applied in areas such as inconsistency handling in OWL ontologies and ontology mapping [17]. These areas start from the hypothesis that ontologies do not handle uncertainty, but produce uncertainty when grouping together a set of ontologies.

Recently a new approach named BeliefOWL [18] has been proposed, which does not deal with ontology mapping or inconsistencies; it focuses rather on translating an OWL taxonomy into a directed evidential network based on the evidence theory. Figure 7 below illustrates the framework of BeliefOWL which translates OWL ontology into DAG named Directed Evidential Network (DEVN). Further discussion regarding the details of this framework can be seen in [18].

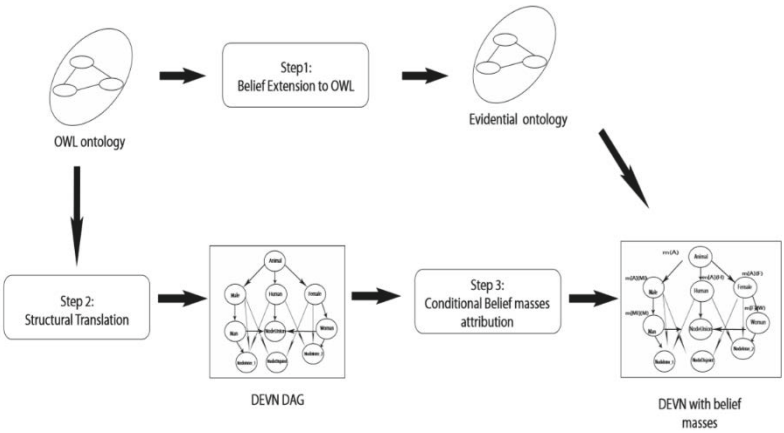


Fig. 7. BeliefOWL framework (*Source*: Essaid, 2009)

3 Methodology

The methodology used in this study refers to the following stages:

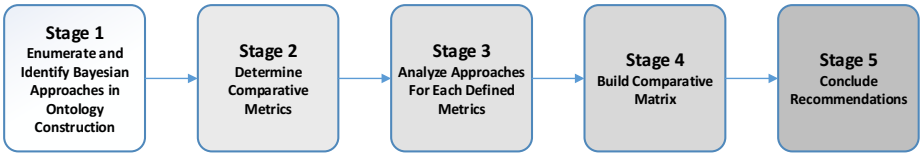


Fig. 8. Method used to conduct the comparative study

Stage 1: Enumerate and Identify Bayesian Approaches in Ontology Construction. At this stage, study of literature carried out to look for approaches that have been proposed by previous researchers to deal with the uncertainty in ontology. Understanding of each approach in detail needs to be done to see the differences. A similar approach to the other will be considered as the same approach. The output of the process at this stage is a list of Bayesian-based approaches to deal with the uncertainty in ontology.

Stage 2: Determine Comparative Metrics. At Stage 2, we determine variables as metrics that will be examined into every approach that has been obtained in Stage 1. The chosen variables must be able to demonstrate the possibility or ease of implementation, or to reflect the performance of an approach.

Stage 3: Analyze Approaches for Each Defined Variables. After comparative variables defined, the next step is to examine these variables into approaches. Each variable in the approaches will be assessed.

Stage 4: Build Comparative Matrix. The next step is to summarize the results of Stage 3, then build a comparative matrix of approaches-variables.

Stage 5: Conclude Recommendations. Based on results of Stage 3 and 4, we build the recommendations as a reference in using approaches to fit the case at hand. Recommendations can be in the form of a determination of the most appropriate approach to choose, because it has the highest scores for most or all of the variables tested.

4 Implementation

4.1 Stage 1: Enumerate and Identify Bayesian Approaches in Ontology Construction

Articles related to the use of Bayesian in the ontology has been searched using keywords: BAYESIAN ONTOLOGY in the journal Elsevier, IEEE, and SpringerLink. The results are grouped into four approaches: BayesOWL or OntoBayes, Multi-Entity Bayesian Networks (MEBN), Probabilistic OWL (PR-OWL), and the Dempster-Shafer Theory.

4.2 Stage 2: Determine Comparative Metrics

We choose variables which are features of the BN for comparison. Variables that could be reflected specific characteristics possessed by each approach. The comparative variables that we choose to examine are:

1. *Complexity*. Describing the complexity of the approach. The more numerous and severe measures should be taken to achieve the objectives, it means that the approach is more complex. The values given are SIMPLE, MODERATE, or COMPLEX.
2. *Accuracy*. Describing how close the result obtained with the set target. The closer the distance, the higher the level of accuracy. The values given are LOW, MEDIUM, or HIGH.
3. *Ease of Implementation*. The degree of ease of the approach to be applied technically. The values given are EASY, MEDIUM, or HARD.
4. *Reasoning mechanism*. The assessment of whether there is a mechanism to explain the reasoning used so that it can be traced. The values given are AVAILABLE or UNAVAILABLE.
5. *Tools support*. Describes whether the modeling software to support the implementation of these approaches are available or not. The values given are UNAVAILABLE, AVAILABLE BUT INSUFFICIENT, or AVAILABLE AND SUFFICIENT.

4.3 Stage 3: Analyze Approaches for each Defined Metrics

BayesOWL provide a powerful framework but still simple to model the probability in ontology. Powerful because it can model the probability using five basic relation functions: intersection, union, complement, disjoint, and equivalent. Simple, because it is done by adding nodes that have been provided by the framework to construct a whole ontology.

BayesOWL was proposed to model uncertainty in OWL ontologies through BN by adding a second ontology which declares the probabilistic relationships. The advantage is that neither OWL nor ontologies defined in OWL need to be modified [4]. Reasoning in BayesOWL achieved by using Bayesian inference mechanism. Therefore, BayesOWL ontology must be first converted into a form of Bayesian Network.

BayesOWL implementation can be done in several ways. First, users can directly implement it using OWL editor (e.g. Protégé [21]) and create its' control nodes manually. Using this way, it is necessary for the user to know the concept of using the BayesOWL framework to build DAG. Second, by using BayesOWL executable application that has been built equipped with a Graphical User Interface (GUI). User input files of ontology and probability in .owl format, then the application will process both of files and build CPT from translated BN. The third way is to use the Java API in the form of packages. There are six available packages where each package can be used separately, such as package `coreAlgorithms`, `parser`, and `GUI`. Further discussion about the technical implementation of BayesOWL can be found in [19].

MEBN offers a complex framework for implementing the BN to support the application of probability to the structure of the ontology. There are some features of MEBN (most notably context constraints) that cannot be represented declaratively in standard relational languages. MEBN should be combined with other modeling languages such as Quiddity*Suite [15] based on Prolog-style rules that provides the ability to specify a much more powerful declarative representations.

Constructing MFrags and all their elements to support probabilistic ontology is also have to be done manually while the whole process is error-prone and tedious. Avoiding errors or inconsistencies requires learning very deep knowledge of the logic and of the data structure of MEBN/PR-OWL [11]. In order to implement it, one has to import the PR-OWL ontology into an ontology editor (e.g. Protégé), then one can start constructing domain-specific concepts using the PR-OWL definitions to represent uncertainty about their attributes and relationships according to the Multi-Entity Bayesian Network (MEBN) model.

A graphical user interface (UnBBayes-MEBN) is available to make it easier to instantiate the main ontological concepts (M-Frags, Resident, Input, Context Nodes, etc). UnBBayes-MEBN implements also a reasoner based on the PR-OWL/MEBN framework. Unfortunately, this GUI has a General Public License (GPL) which is a little too restrictive in order to integrate it into a commercial application. In addition to our knowledge, MEBN community is not wide enough to be considered as an emerging standard. Thus it represents a major difficulty to manipulate this tool.

PR-OWL is an upper ontology to guide the development of a probabilistic ontology developed based on MEBN framework. PR-OWL for probabilistic system consists of a set of classes, subclasses and properties that collectively form a framework for building probabilistic ontologies. The first step to build a probabilistic ontology is by using any OWL editor to import an OWL file containing PR-OWL classes, subclasses, and properties.

There is an important difficulty when trying these tools in regards to the type of knowledge available. It is because PR-OWL does not have any procedure to check the consistency of CPT (such as D-IPFP in BayesOWL). As a matter of fact, the representation of information absence is badly taken into account by the theory of probability. This requirement often lead in using a symmetry (minimax error) argument to

assign prior probabilities to random variables (e.g. assigning 0.5 to values for which no information is available).

A PR-OWL plugin for current OWL ontology editors is a priority for future efforts [9]. The process of writing probabilistic ontologies can be greatly improved via automation of most of the steps in the ontology building, not only for defining MFrags to represent sets of related hypotheses, but also for consistency checking, reified relations and other tasks that demand unnecessary awareness of the inner workings of the present solution. Once implemented, such a plugin has the potential to make probabilistic ontologies a natural, powerful tool for the Semantic Web.

Dempster-Shafer theory uses a somewhat different approach than the others, because it is based on "belief" rather than "probable". One of the major advantages of Dempster-Shafer theory over probability theory is thus to allow one to specify a degree of inaccuracies or ignorance in a situation instead of being forced to supply prior probabilities, although we can provide it as well. Incomplete information in priors and conditionals can be used as a support by calculating it using Bayes equations to produce values of degree of ignorance.

The reasoning mechanism of Dempster-Shafer framework is performed through three steps. The first step, one should use a published reasoning engine (e.g. Jena Framework [23]) that provides programmatic environment for RDF, RDFS, OWL, and SPARQL, including a rule-based inference engine. The second stage is transmit the query results to a Dempster-Shafer Library [24] that provides basic functions for calculating the mass, credibility, and plausibility. The third stage, the output of Dempster-Shafer running model are then juxtaposed with the OWL source for analysis as the process of reasoning. Unfortunately, those three stages still to be done separately, no tools are capable implementing it into an environment.

4.4 Stage 4: Build Comparative Matrix

Analysis for each approach on the tested metrics gives an overview of how the nature, features, and performance of each of these approaches. Table 1 below summarize the analytical results.

Table 1. The comparative matrix of Approaches-Metrics

Metrics / Approaches	Complexity	Accuracy	Ease of Implementation	Reasoning Mechanism	Tools Support
BayesOWL	SIMPLE/ MODERATE	MEDIUM	EASY	AVAILABLE	AVAILABLE AND SUFFICIENT
MEBN	COMPLEX	HIGH	MEDIUM/HARD	AVAILABLE	AVAILABLE BUT INSUFFICIENT
PR-OWL	COMPLEX	HIGH	MEDIUM/HARD	AVAILABLE	AVAILABLE BUT INSUFFICIENT
Dempster-Shafer	MODERATE	MEDIUM	MEDIUM	AVAILABLE	UNAVAILABLE

4.5 Stage 5: Conclude Recommendations

Each approach has been analyzed and the summary of the results are presented in Table 1 above. From the table we can see that the whole approach has provided a reasoning mechanism. MEBN and PR-OWL has the same value for all tested variables. This is reasonable because the PR-OWL itself is an enhancement of MEBN. Dempster-Shafer promises a lot of interesting things in the handling of uncertainty in the ontology with a level of complexity that remains moderate, but supporting tools for the implementation of the Dempster-Shafer still not yet available. BayesOWL provides a good approximation, simpler than the PR-OWL, easy to implement, and has adequate tools support. BayesOWL or OntoBayes is the most recommended approach as an extension of OWL to handle uncertainty in the ontology construction.

5 Conclusion

Bayesian-based approaches to deal with uncertainty in ontology construction have been proposed. BayesOWL, one of the proposed approach has advantages compared to other approaches and most likely to be implemented with a relatively less effort. BayesOWL can be used to denote the degree of certainty of the relationship between the resources in the ontology. The interest thing to be explored in future work is how to choose the relationship between a resources to other resources in an ontology and determine its' degree of certainty. If this can be achieved, then the system can perform automatic resource property addition and selection in ontology construction.

References

1. Santos Jr., E., Jurmain, J.C.: Bayesian Knowledge-driven Ontologies: Intuitive Uncertainty Reasoning for Semantic Networks. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 856–863 (2011)
2. Ding, Z., Peng, Y., Pan, R., Yu, Y.: A Bayesian Methodology towards Automatic Ontology Mapping. In: Proceedings of the AAAI-05 C&O Workshop on Contexts and Ontologies: Theory. AAAI Press (2005)
3. Ding, Z., Peng, Y., Pan, R.: A Bayesian Approach to Uncertainty Modelling in OWL Ontology. In: Proceedings of the International Conference on Advances in Intelligent Systems - Theory and Applications, Luxemburg (2004)
4. Bellenger, A., Gatepaille, S.: Uncertainty in Ontologies: Dempster-Shafer Theory for Data Fusion Applications. In: Workshop on Theory of Belief Functions, Brest, France (2010)
5. Noy, N.F., Rector, A.: Defining N-ary relations on the Semantic Web: Use with individuals, in W3C Working Draft, World Wide Web Consortium: Boston, MA. p. W3C Working Draft (2004)
6. Perkusich, M., Soares, G., Almeida, H., Perkusich, A.: A Procedure to Detect Problems of Processes in Software Development Projects using Bayesian Network. Expert Systems with Application 42, 437–450 (2015)
7. Laskey, K.B.: MEBN: A Language for First-Order Bayesian Knowledge Bases. Artificial Intelligence 172(2-3), 140–178 (2008)

8. da Costa, P.C.G., Laskey, K.B.: Multi-Entity Bayesian Networks Without Multi-Tears. Research Draft. Department of Systems Engineering and Operation Research, George Mason University, Fairfax, USA (2005)
9. Park, C.Y., Laskey, K.B., Costa, P., Matsumoto, S.: Multi-Entity Bayesian Networks Learning In Predictive Situation Awareness. In: 18th International Command And Control Research And Technology Symposium, Alexandria, Virginia USA (2013)
10. da Costa, P.C.G., Laskey, K.B.: MEBN Fragments, http://www.pr-owl.org/mebn/mfrags.php (accessed on December 17, 2014)
11. da Costa, P.C.G., Laskey, K.B., Laskey, K.J.: PR-OWL: A Bayesian Ontology Language for the Semantic Web. In: da Costa, P.C.G., d'Amato, C., Fanizzi, N., Laskey, K.B., Laskey, K.J., Lukasiewicz, T., Nickles, M., Pool, M. (eds.) URSW 2005 - 2007. LNCS (LNAI), vol. 5327, pp. 88–107. Springer, Heidelberg (2008)
12. Costa, P., Laskey, K.B.: PR-OWL: A Framework for Probabilistic Ontologies. In: Proceedings of the Conference on Formal Ontologies and Information Systems, Amsterdam, The Netherlands, pp. 237–249 (2006)
13. Bedini, I., Nguyen, B.: Automatic Ontology Generation: State of the Art. University of Versailles Technical Report (2007)
14. Carvalho, R.N., Laskey, K.B., Costa, P.: PR-OWL 2.0 Bridging the Gap to OWL Semantics. In: Proceedings of the Sixth International Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2010), held at the Nineth International Semantic Web Conference (ISWC 2010), Washington DC, USA (2009)
15. Laskey, K.B., Haberlin, R., Carvalho, R.N., Costa, P.: PR-OWL 2 Case Study: A Maritime Domain Probabilistic Ontology. In: Proceedings of the 6th International Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS 2011), Fairfax, VA, USA (2011)
16. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
17. Yaghlane, B.B., Laamari, N.: OWL-CM: OWL Combining Matcher based on Belief Functions Theory. In: Proceedings of the 2nd International Workshop on Ontology Matching (2007)
18. Essaid, A., Yaghlane, B.B.: BeliefOWL: An Evidential Representation in OWL Ontology. In: International Semantic Web Conference, International Workshop on Uncertainty Reasoning for the Semantic Web, Washington DC, USA, pp. 77–80 (2009)
19. Zhang, S., Sun, Y., Peng, Y., Wang, X.: BayesOWL: A Prototype System for Uncertainty in Semantic Web. In: Proceedings of the 2009 International Conference on Artificial Intelligence (ICAI), Las Vegas Nevada, USA (2009)
20. IET. Quiddity Suite: Overview, http://www.iet.webfactional.com/quiddity.html (accessed on December 17, 2014)
21. Stanford. Protégé, http://protege.stanford.edu/ (accessed on January 24, 2015)
22. Daconta, M.C., Obrst, L.J., Smith, K.T.: The Sematic Web: A Guide to the Future of XML, Web Services, and Knowledge Management, p. 312. Wiley Publishing, Inc., Indianapolis (2003)
23. Apache. Apache Jena, https://jena.apache.org/ (accessed on January 24, 2015)
24. Reineking, T.: Open Source Java Dempster-Shafer Library, http://www.softpedia.com/get/Programming/Components-Libraries/Java-Dempster-Shafer.shtml (accessed on January 24, 2015)

Applicability of Cyclomatic Complexity on WSDL

Sanjay Misra[1], Luis Fernandez-Sanz[2], Adewole Adewumi[1],
Broderick Crawford[1,2,3], and Ricardo Soto[1,4,5]

[1] Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
[2] Universidad Finis Terrae, Santiago, Chile
[3] Universidad San Sebastián, Santiago, Chile
[4] Universidad Autónoma de Chile, Santiago, Chile
[5] Universidad Central de Chile, Santiago, Chile
{sanjay.misra,wole.adewumi}@covenantuniversity.edu.ng,
luis.fernandezs@uah.es,
{broderick.crawford,ricardo.soto}@ucv.cl

Abstract. Complexity metrics are useful for predicting the quality of software systems because they quantify the quality attributes. Web services, a new kind of software system has been providing a common standard mechanism for interoperable integration of disparate systems and gaining a great deal of acceptance by different types of parties that are connected to the internet for different purposes. In this respect, quality of the web-services should be quantified for easy maintenance and quality of services. Further, the Web Services Description Language (WSDL) forms the basis for Web Services. In this paper, we are evaluating the quality of the WSDL documents by applying the Cyclomatic Complexity metric, a well known and effective complexity metric, which has not been used to evaluate the quality of WSDL till date.

Keywords: WSDL, web-services, software metrics, cyclomatic complexity.

1 Introduction

Software metrics have been the means through which software engineers have tried to ensure quality. However, absolute measures are rare in software engineering [9]. Instead, software engineers attempt to derive a set of indirect measures that lead to metrics that provide an indication of quality of some representation of software. Software engineers plan 'how' an information system should be developed in order to achieve its quality objectives. The quality objectives may include: maintainability, availability, reliability and performance [10], which are closely related to software complexity.

The researches on web services are in developing stage. The existing standards are under revisions and new technologies are evolving. Web services can be treated as a software component, whose data can be presented using XML Schema, and their interfaces through Web Service Definition Language (WSDL [1]). Web Services Description Language (WSDL) is an eXtensible Markup Language (XML) based document and also a language which continuously gaining popularity for web services. Since WSDL is a language, therefore its quality should be quantified.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 247–254, 2015.
DOI: 10.1007/978-3-662-46742-8_23

Software metrics always play an important role in developing software. In addition, their importance increases as companies and organizations begin adopting these new technologies. In recent years, companies such as IBM, Microsoft, BEA Oracle and Borland have started adopting Web-services - a new type of application [3]. However, the quality of web services is not been evaluated through the applicability of very well known software metrics like cyclomatic complexity, Halsted programming efforts. Cyclomatic complexity is one of the most popular and a well-accepted complexity metric, which is in use since the last three decades.

The degree of a software project's complexity impacts on its maintainability. The complexity degree of a Web service can thus be measured by analyzing its WSDL [4], [5], [6], [7] document because WSDL provides the description of the Web service to the service requestor's software development [8].

Actually, our motivation for the applicability of cyclomatic complexity arises from the paper of Yijun et al. [3]. They argued that since there are no implementation details of WSDL, we cannot measure the metrics that are based on the complete source code, such as McCabe complexity [2] that measures the controls flow. However, in this paper we are showing that the cyclomatic complexity can be used to evaluate the WSDL.

2 Web Services and WSDL

The following figure shows how WSDL is used in web services.

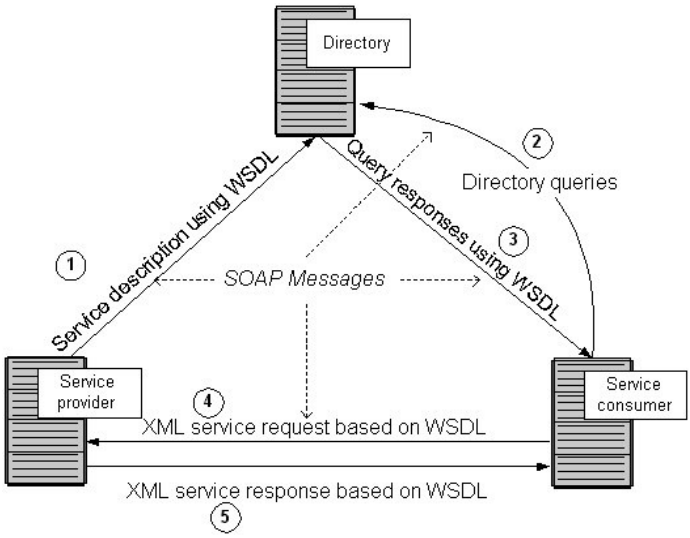


Fig. 1. This figure shows the use of WSDL in Web Services [11]

WSDL is the foundation layer of Web Services. We can easily observe from the figure that how at the two ends, service provider and service consumers are communicating through WSDL.

3 Cyclomatic Complexity

McCabe [2] devised a graph theoretic complexity measure for managing and controlling program complexity. The metric depends on the decision structure of a program and not its physical size and as a result is calculated from its flow graph representation. In addition, if a program has a number of modules then its complexity is calculated as the sum of cyclomatic complexity of the modules calculated individually. This measure finds the number of linearly independent circuits that are required to provide information about all circuits and thus attempts, to determine the number of execution paths in a program. The number of fundamental circuit also acts as an index of program testing effort. This measure is most widely used one amongst the different available measure for control flow complexity.

The cyclomatic complexity of a part of code is the count of the number of linearly independent paths through the code. For example, if the code contained no decision points such as 'for' loops, the complexity would be only 1, as there is only one path through the code.

The cyclomatic *complexity* of a structured program is defined with reference to a directed graph containing the basic blocks of the program, with an edge between two basic blocks if control may pass from the first to the second (the *control flow graph* of the program). The complexity is then defined as:

$$CC = E - N + 2P \tag{1}$$

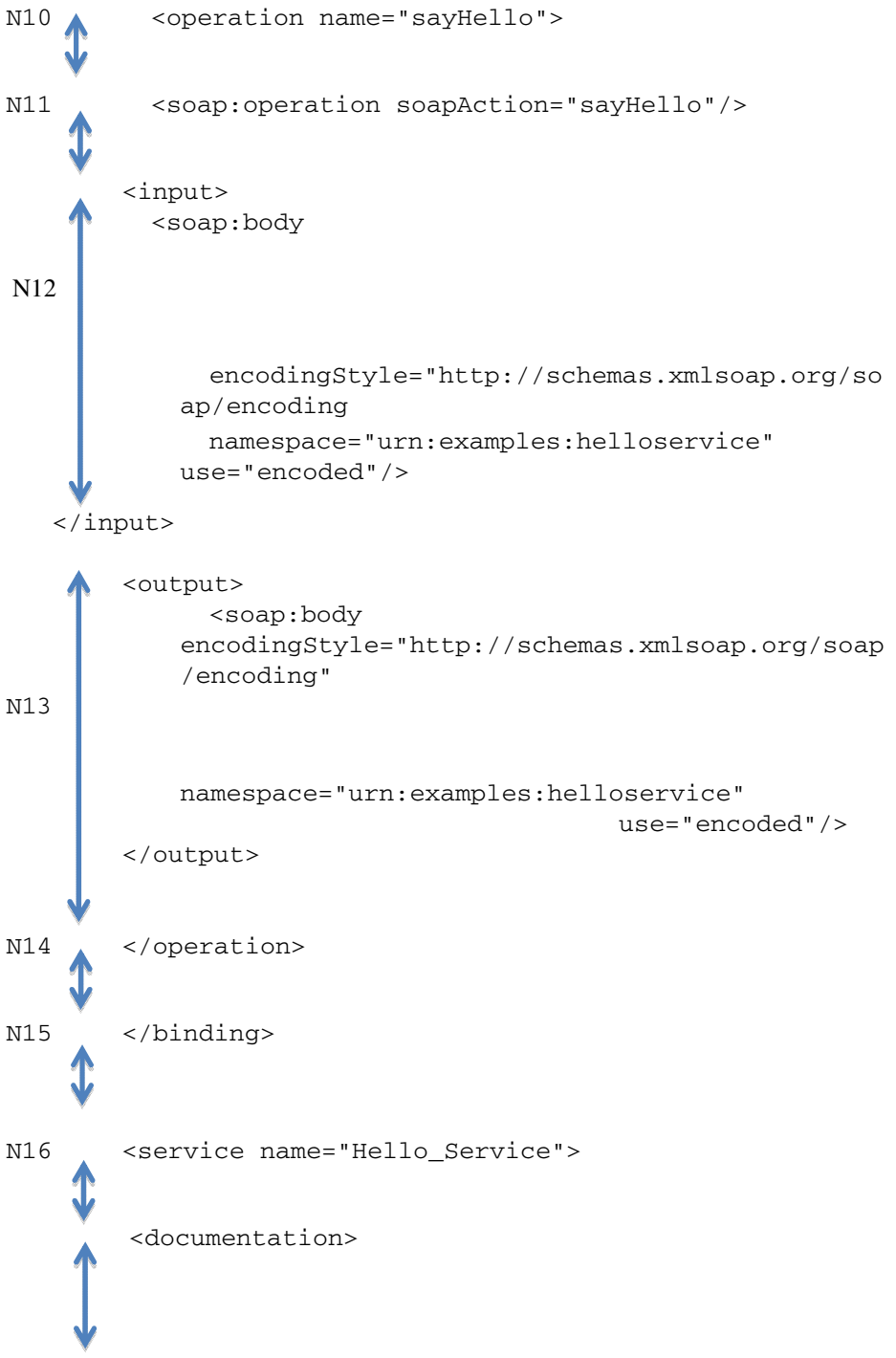
Where,
CC = Cyclomatic complexity
E = the number of edges of the graph
N = the number of nodes of the graph
P = the number of connected components.

4 Demonstration of CC for WSDL

4.1 Demonstration of Nodes in WSDL

```
<?xml version="1.0" encoding="UTF-8"?>
          <definitions name="HelloService"
        targetNamespace=http://www.ecerami.com/wsdl/He
        lloService.wsdl
        xmlns=http://schemas.xmlsoap.org/wsdl/
```

```
N0      xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
          xmlns:tns="http://www.ecerami.com/wsdl/Hello
        Service.wsdl"
        xmlns:xsd=http://www.w3.org/2001/XMLSchema >
      <message name="SayHelloRequest">
N1     <part name="firstName" type="xsd:string"/>
      </message>

      <message name="SayHelloResponse">
N2     <part name="greeting" type="xsd:string"/>
      </message>


N3              <portType name="Hello_PortType">

N4        <operation name="sayHello">

N5      <input message="tns:SayHelloRequest"/>

N6        <outputmessage="tns:SayHelloResponse"/>

 N7       </operation>

 N8        </portType>


          <binding name="Hello_Binding"

          type="tns:Hello_PortType">
N9         <soap:binding style="rpc"

     transport="http://schemas.xmlsoap.org/soap/http"/>
```


N10
<operation name="sayHello">
N11
<soap:operation soapAction="sayHello"/>
<input>
<soap:body
N12
encodingStyle="http://schemas.xmlsoap.org/soap/encoding
namespace="urn:examples:helloservice"
use="encoded"/>
</input>
<output>
<soap:body
encodingStyle="http://schemas.xmlsoap.org/soap/encoding"
N13
namespace="urn:examples:helloservice"
use="encoded"/>
</output>
N14
</operation>
N15
</binding>
N16
<service name="Hello_Service">
<documentation>

```
N17     WSDL File for HelloService

         </documentation>

              <port binding="tns:Hello_Binding"
            name="Hello_Port">
N18     <soap:address
              location="http://localhost:8080/soap/servlet
            /rpcrouter"/>
          </port>

 N19     </service>

N20     </definitions>
```

4.2 Graph Representation of the WSDL

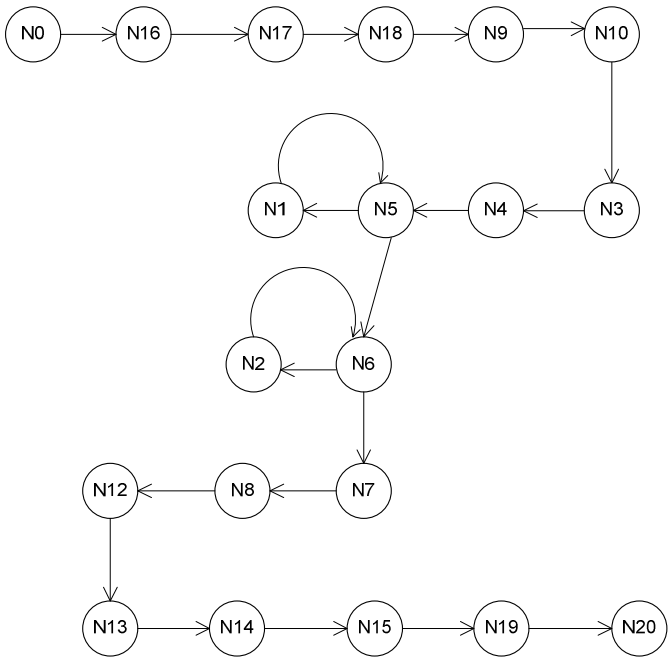


Fig. 2. This figure shows the control flow diagram for WSDL document

The Cyclomatic Complexity of the given WSDL is given by:

$CC = E - N + 2P$

$CC = 21 - 20 + 2*1$

$CC = 3$

Table 1. Recommended thresholds for Cyclomatic complexity

Cyclomatic Complexity Values	Risk
1 - 10	Low Risk Program
11 - 20	Moderate Risk
21 - 50	High Risk
>50	Most complex and Highly Unstable method

5 Validation

Weyuker proposed nine properties to make judgment about any proposed software complexity measure [12]. These properties evaluate the weaknesses of a proposed measure in a practical way. By their help, one may obtain an idea about the validity of his/her own proposal. In fact, Weyuker herself evaluated the four complexity measures against these properties. Cyclomatic complexity was one of the complexity measures, which was evaluated by Weyuker's properties. Cyclomatic complexity was satisfied by six Weyuker's properties. Actually, the cyclomatic complexity, which was evaluated by Weyuker's properties, was basically for procedural languages. But all the arguments, which were given for the validation of cyclomatic complexity for procedural language, are also valid for the proof of CC for WSDL. In this respect, we can conclude that, our measure is also satisfied by six Weyuker's properties. Here it is worth mentioning that it is not required for any complexity measure to satisfy all Weyuker's properties. This is because some of the Weyuker's properties are contradictory in nature. Therefore, at most six or seven Weyuker's properties satisfy the complexity measures.

6 Conclusion

Evaluating quality of WSDL through software metrics is a new area of research. Since there has been little number of researches so far, adaptation of some existing software metrics to the WSDL is helpful for the evaluation of the quality of WSDL. It is suggested that Cyclomatic complexity metric can be considered in the design, development and operational phase of the web-services.

References

1. W3C, Web Services Description Language (WSDL) 1.1 (2001)
2. McCabe, T.: A Complexity Measure. IEEE Trans. Software Eng. 2, 308–320 (1976)
3. Yu, Y., Lu, J., Fernandez-Ramil, J., Yuan, P.: Comparing Web Services with other Software Components. In: Proc. IEEE International Conference on Web Services, pp. 388–397 (2007)
4. Erl, T.: Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services. Prentice Hall Publishers (2004)
5. Newcomer, E., Lomow, G.: Understanding SOA with Web Services. Addison Wesley Professional (2004)
6. Cerami, E.: Web Services Essentials, Distributed Applications with XML-RPC, SOAP, UDDI & WSDL. O'Reilly Publishers (2002)
7. Weerawarana, S., Curbera, F., Leymann, F., Storey, T., Ferguson, D.F.: Web Services Platform Architecture: SOAP, WSDL, WS-Policy, WS-Addressing, WS-BPEL, WS-Reliable Messaging, and More. Prentice Hall Publishers (2005)
8. http://www.w3.org/TR/2004/REC-xmlschema-1-20041028/
9. Pressman, R.S.: Software Engineering: A Practitioner's approach, 5th edn. McGraw-Hill (2001)
10. Sommerville, I.: Software Engineering, 6th edn. Addison-Wesley (2001)
11. http://www.service-architecture.com/web-services/articles/web_services_explained.html
12. Weyuker, E.J.: Evaluating software complexity measure. IEEE Transaction on Software Complexity Measure 14, 1357–1365 (1988)

Feature Extraction

Multiclass Fruit Classification of RGB-D Images Using Color and Texture Feature

Ema Rachmawati, Iping Supriana, and Masayu Leylia Khodra

School of Electrical Engineering and Informatics
Bandung Institute of Technology, Indonesia
ema.rachmawati22@students.itb.ac.id,
{iping,masayu}@stei.itb.ac.id

Abstract. Fruit classification under varying pose is still a complicated task due to various properties of numerous types of fruit. In this paper we propose fruit classification method with a novel descriptor as a combination of color and texture feature. Color feature is extracted from segmented fruit image using Color Layout Descriptor, while texture feature is extracted using Edge Histogram Descriptor. Support Vector Machine (SVM) with linear and RBF kernel is used as classifier with 10-fold cross validation. The experimental results demonstrated that our descriptor achieves classification accuracy of over 93.09 % for fruit subcategory and 100 % for fruit category from over 4200 images in varying pose.

Keywords: fruit classification, color; texture, SVM, Color Layout Descriptor, Edge Histogram Descriptor.

1 Introduction

Object recognition is a necessary stage in the application of computer vision to a higher level, such as activity recognition, full scene content understanding, and so on. Even object recognition system also supports a wide range of practical applications, such as, intelligent video surveillance, robotic navigation, content-based image retrieval, image-based rendering, photo manipulation, augmented reality, fruit recognition, etc. [1]. A major challenge in visual object recognition system is divided into two groups, which are related to system robustness and computational complexity and scalability. Belong to the first group is a challenge in handling intra-class variations in appearance (different appearance from a number of objects of the same category) and inter-class variations [2]. Instances of the same object category can generate different images caused by a variety of variables that influence, such as illumination, object pose, camera viewpoint, partial occlusion, and background clutter. While the challenges which belong to the second group were i.e. very large objects of different categories, high-dimensional descriptors (which represent the extracted features), difficulties in obtaining labeled training samples without any ambiguity, etc. [1].

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 257–268, 2015.
DOI: 10.1007/978-3-662-46742-8_24

Some research on fruit classification has been done by [3],[4],[5],[6]. Arivazhagan et. al [3] make a fruit recognition system using a dataset consists of 15 fruit categories fruit. Seng & Mirisaee [4] used 6 fruit categories (apple, strawberry, durian, watermelon, banana, and orange) with number of instances about 50 images. Kavdir & Guyer [5] conducted fruit classification of *Golden Delicious Apple* using 181 apples in their experiment. While [6] used 15 categories of fruit (Plum, Agata potato, potato asterix, cashew, onion, orange, Taiti lime, kiwi, Fuji apple, granny-smith apple, watermelon, honeydew melon, nectarine, Williams pear, diamond peach) in their experiment.

Contribution and Overview. In this paper, we introduce an efficient feature extraction algorithm and use it for recognizing fruit under varying pose. Until now, fruit recognition under varying pose is still a complicated task due to various properties of numerous types of fruit. Paper [3],[4],[5],[6] have not considered pose variation of fruit. We propose a method to classify fruit with some varying pose using color and texture feature of MPEG-7 visual descriptor [7] to build feature representation effectively. Fruit dataset were captured from a turntable with a mounted camera on three different height above horizon [8] while were rotated along with the turntable, in order to get the variations of fruit pose.

The remainder of the paper is organized as follows. Section 2 gives an overview of the related work. In Section 3 we present the method we use in feature extraction. We present the experimental results in Section 4. Finally, we conclude the paper in Section 5.

2 Related Work

MPEG-7 visual description tools contains basic structure and descriptor that includes basic visual features: color, texture, shape, motion, and localization [9]. Among the four color descriptors exist in MPEG-7, Scalable Color Descriptor (SCD) and Color Layout Descriptor (CLD) were color descriptor that were often used because they are easier to be implemented and deployed [7], [10], [11],[12]. SCD tried to improve the performance of the use of color histogram as a descriptor by implementing Haar Transform, while CLD used Discrete Cosine Transform on the YCbCr color space.

Scalable Color Descriptor (SCD) as one of MPEG-7 visual descriptor is a histogram-based descriptors that require maximum 1024 bits per histogram [7], [13],[14]. The maximum number of bin is 256 bins, which will gradually be reduced using Haar Transform. Such down scaling affects the retrieval performance significantly. The use of 16 or 32 bits should be enough to achieve good performance [13]. Reference [15] used some MPEG-7 visual descriptor to evaluate retrieval effectiveness for searching in large databases.

As stated in [16], Color Layout Descriptor was recommended as one of good color descriptor besides Dominant Color Descriptor [17]. CLD is usually used in image retrieval systems, such as in [10], [11],[18]. Reference [10] used color layout descriptor as a feature description for high-speed image/video segment retrieval. The image retrieval system introduced by [18] is based on a query by layout method using CLD and EHD. While [11] combine CLD with texture descriptor (Gabor filters) to con-

struct robust feature set in CBIR system. We will use those three descriptor (SCD, CLD, and EHD) as they are widely used in various literature [15], for building feature representation and using it in our fruit recognition system.

3 Feature Extraction

3.1 Color Feature Extraction

Spatial distribution of color features can be represented by Color Layout Descriptor. It captures the spatial layout of the representative colors on a region based on DCT and expresses in the YcbCr color space Input image is first divided into 64 non-overlapping blocks. Then representative color is determined from each block. By applying the 8-point Discrete Cosine Transform (DCT) after converting color space from RGB into YCbCr and performing non-linear quantization of the zigzag scanned DCT coefficients, the CLD is obtained. This process can be seen in Fig. 1.

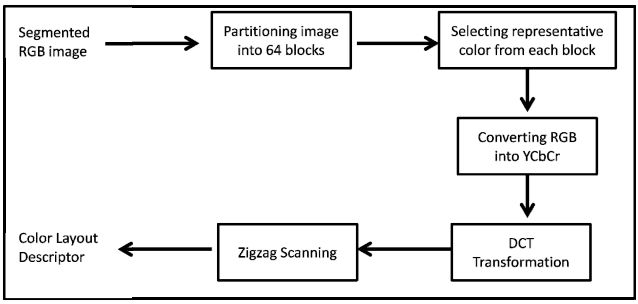


Fig. 1. CLD block diagram [11]

3.2 Texture Feature Extraction

Texture is set of metrics calculated in image processing designed to quantify the perceived texture of an image. Image Texture gives us information about the spatial arrangement of color or intensities in an image or selected region of an image [19]. Edge Histogram Descriptor (EHD) that is used in [20] to represent the texture features describes the spatial distribution of five types of local edge i.e., vertical, horizontal, 45°, 135°, and undirected (Fig. 2). Image is divided into 16 blocks, and then 5 bins histogram generated from each block to obtain an 80-dimensional feature vectors.

4 Experiment and Results

Block diagram of multiclass fruit classification can be seen in Fig. 3. The extraction of color and texture feature is conducted in order to build feature representation. Later on, the learning and classifier evaluation are conducted on the generated feature vectors using supervised learning methods. Color feature extraction was conducted using Color Layout Descriptor (CLD) and texture extraction using Edge Histogram Descriptor (EHD).

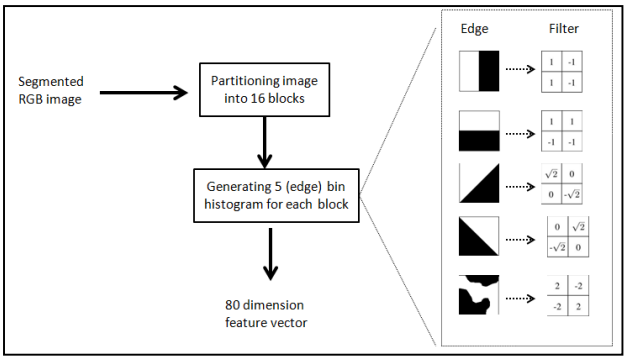


Fig. 2. EHD block diagram

We conducted experiment in order to analyze classification performance based on color and texture. Dataset and method used in this experiment were explained in section 4.1. We analyzed our system ability to classify each image into corresponding fruit categories and subcategories. In category classification we classify fruit into seven fruit categories, while in the subcategory classification we classify fruit into 32 subcategories.

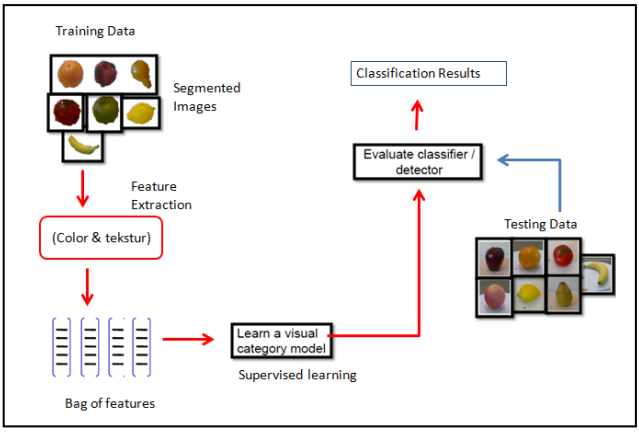


Fig. 3. Block diagram of multiclass fruit classification

We perform classification using MPEG-7 color and texture descriptor [13], that is CLD and EHD and compare with SCD as they are widely used in various literature [15]. We compare classification accuracy of using color feature only and of using color and texture feature combination. We also measure the time required for the training process (CPU Time Training) and the size of the feature vector generated on the feature extraction process as a mechanism for evaluating the system.

4.1 Dataset for Training and Testing

The dataset used in the experiment is fruit images in RGB-D Object dataset [8]. The fruit was placed on a turntable and captured with a camera mounted at three different heights relative to turntable, about 30°, 45°, and 60° above the horizon [8]. Some samples of fruit images can be seen in Fig. 4. Number of fruit image used in this experiment is 4290 images and each fruit has some fruit subcategories. Each subcategory consists of 122 – 153 RGB images as shown in Fig. 5. We use SVM as classifier (with linear and RBF kernel) with 10-fold cross validation on Weka [21] and libSVM [22]. The number of training data is 3861 images and testing data is 429 images.

Fig. 4. Samples of Fruit Images from dataset

The number of dimension of extracted feature vector using CLD is 33 dimensions. It is obtained from 21 coefficients of Y component, 6 coefficients of Cb component, and 6 coefficients of Cr component. The number of coefficients Y is greater than the number of coefficients Cb and Cr, because the most important information is stored in component Y than in component Cb and Cr. The number of dimensions of extracted feature of SCD is 32 dimensions. It is the smallest dimension that can be compared to the feature vector of CLD. SCD is a histogram descriptor encoded by Haar Transform and uses the HSV color space uniformly quantized to 256 bins. In this experiment we maintain low dimension of feature vector extracted with SCD in order to avoid large storage size.

4.2 Category Classification

Fig. 6 presents accuracy of 7 fruit category classification using color only features. The highest accuracy obtained using SCD is about 38.65 %, while using CLD is 100 %. Confusion matrix as classification result with feature extracted using SCD (as shown in Table 1) show that classification error occurs in many fruits that have similar colors, such as yellow apple, yellow pear, and banana with lemon. Each column in confusion matrix shows the number of instances in each class of classification results, while each row shows the number of instances in the actual class. For example, in the

Table 1 as many as 235 apple image successfully classified as *apples*, 8 as *banana*, 9 as *pear*, 6 as *peach*, 2 as *lime*, and 355 as a *lemon*.

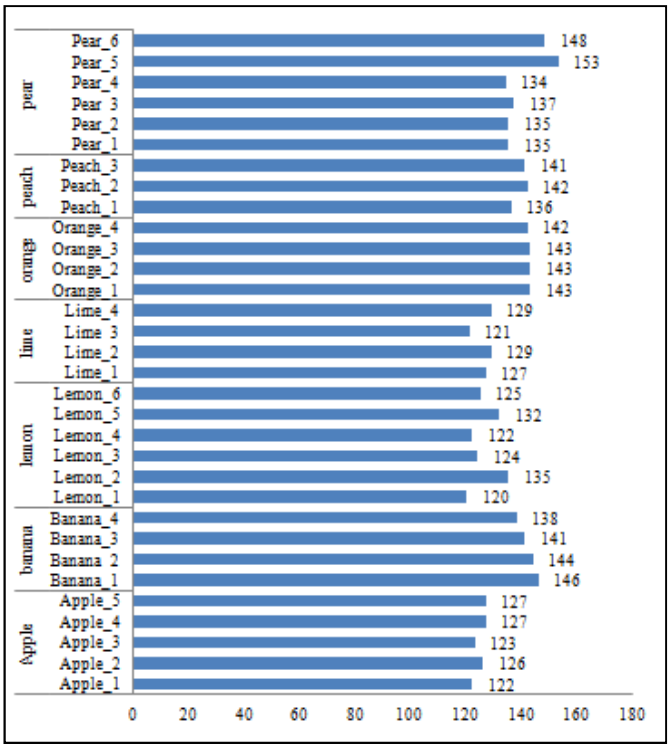


Fig. 5. Number of Images in each subcategory of fruit

Fig. 6. Classification Accuracy of 7 fruit categories using Color Feature

Accuracy of category classification using texture features, EHD, shown in Fig. 7. The combination of SCD and EHD greatly affects the classification accuracy until 88.50 %, as can be seen in Fig. 8. But, involving EHD with CLD does not greatly affect the accuracy of the classification of fruit categories. With linear SVM classifier, accuracy increased by nearly 2% as seen in Fig. 8.

Table 1. SCD – Confusion Matrix for 7 fruit categories classification

Classified as	apple	banana	pear	peach	lime	lemon	orange
apple	**235**	8	19	6	2	355	0
banana	20	**142**	23	1	14	369	0
pear	11	1	**103**	0	11	716	0
peach	39	1	7	**80**	0	292	0
lime	10	6	52	0	**355**	83	0
lemon	1	0	11	1	1	**744**	0
orange	0	0	4	0	0	567	**0**

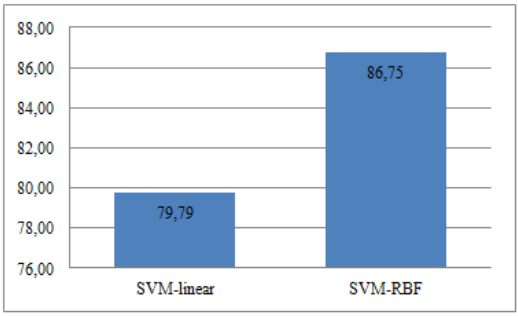


Fig. 7. Classification Accuracy of 7 fruit categories using Texture Feature

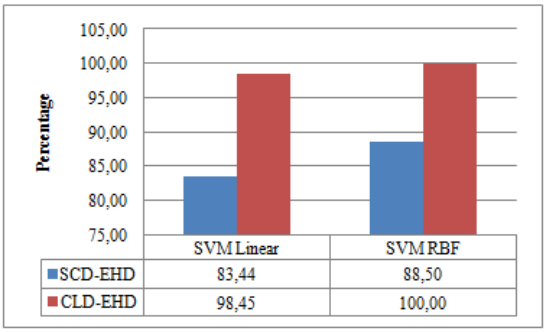


Fig. 8. Classification Accuracy of 7 fruit categories using Color & Texture Feature

4.3 Subcategory Classification

From Fig. 9 we can see that the highest accuracy of subcategory classification with SCD about 19.16 % and with CLD about 91.91 %. Using CLD, there is only an image from subcategory of *banana_2* that misclassified as a subcategory of *pear_5*, because their color is very similar. Other misclassifications are still localized to the same category as can be seen in Fig. 10. For example, *apple_3* successfully classified as *apple_3* as many as 104, and as *apple_4* as many as 19.

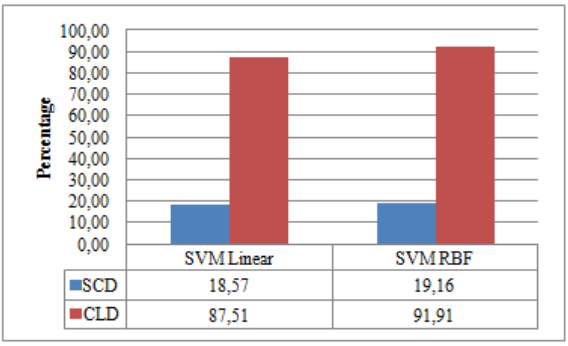


Fig. 9. Classification Accuracy of 32 fruit subcategories using Color Feature

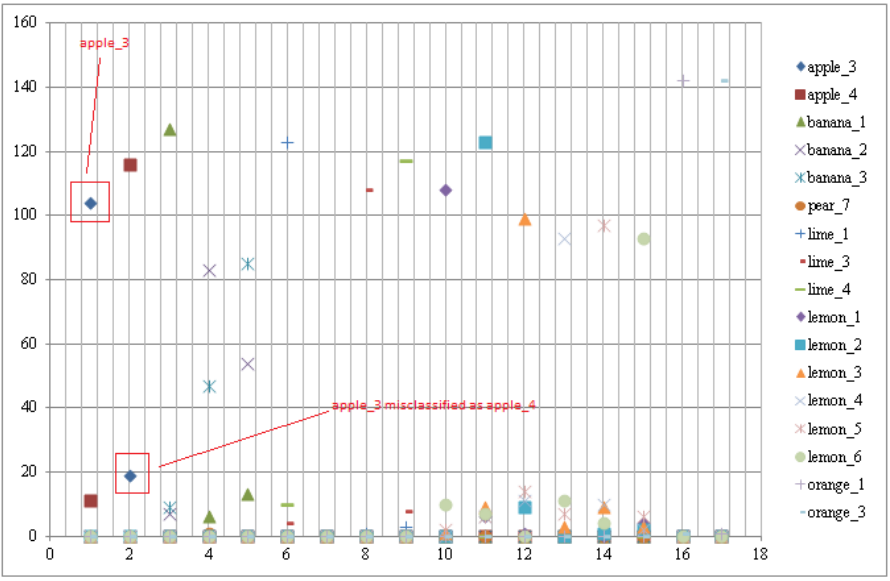


Fig. 10. Scatter chart of 32 Subcategory Classification using CLD

Fig. 11 shows the result of texture descriptor on fruit subcategory classification. It can be seen that the EHD solely cannot achieve satisfying result for subcategory classification task. Thus, we combine color and texture feature into a compact descriptor, with the classification result of combined feature can be seen in Fig. 12.

The combination of SCD and EHD achieves very significant improvement, increased about 50 %. The using of CLD solely can achieve high accuracy result. But the combination of CLD and EHD still achieves a small improvement, about 2 %. Misclassification of an image from subcategory of *banana_2* as subcategory of *pear_5* is no longer occurred, and the number of misclassification in each subcategory become less, as shown in Fig. 13.

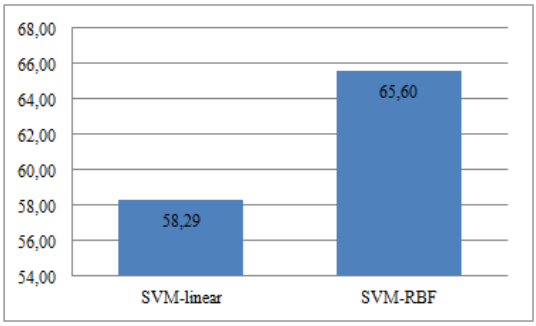


Fig. 11. Classification Accuracy of 32 fruit subcategories using Texture Feature

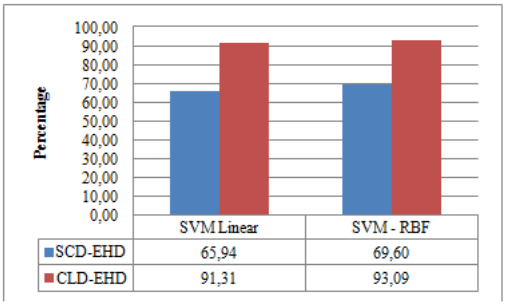


Fig. 12. Classification Accuracy of 32 fruit subcategories Using Color and Texture Feature

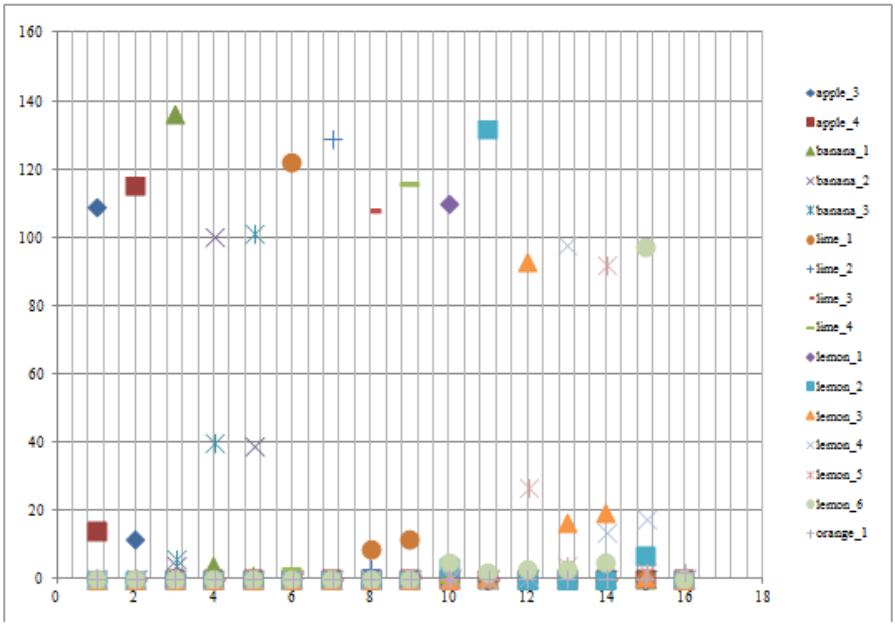


Fig. 13. Scatter chart of 32 Subcategory Classification using CLD & EHD

4.4 CPU Time Training and Storage Size

CPU time training is the time required in the training process of all training data in the dataset. Comparison of CPU Time Training for each descriptor can be seen in Fig. 14. Training time was measured on classification of 32 fruit subcategories. SCD feature vector needs training time longer than CLD or EHD feature vector. The combined feature vector of CLD and EHD needs training time not much different from the combined feature vector of SCD and EHD, about 2 – 4 seconds.

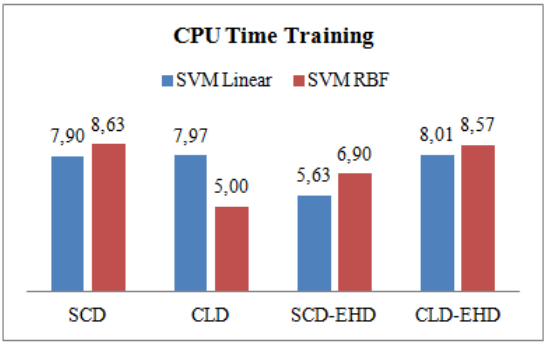


Fig. 14. CPU Time Training for classification of 32 fruit subcategories

The storage size of feature vector extracted from each descriptor can be seen in Table 2. Feature vector extracted by SCD was resulting in smaller storage size than feature vector extracted by CLD because of the quantization process in the resulting histogram.

Table 2. Storage size of each feature vector extracted from each descriptor

Descriptor	Storage size
SCD (32 bin)	0,393 MB
CLD (21 coeff. Y, 6 coeff. Cb, 6 coeff. Cr)	0,733 MB
EHD	0,728 MB
SCD & EHD	1,08 MB
CLD & EHD	1,14 MB

5 Conclusion and Future Work

We presented a combined color and texture feature into a compact descriptor to classify multiclass fruit images with varying pose. The result proves that our combined descriptor is giving high classification rate. Our future work will concentrate on improving classification accuracy with adding shape feature into existing color and texture descriptor. Further, by increasing the number of category and subcategory of fruit images, we hope the classification rate can be improved.

References

1. Zhang, X., Yang, Y.-H., Han, Z., Wang, H., Gao, C.: Object Class Detection: A Survey. ACM Comput. Surv. 46, 10:1–10:53 (2013)
2. Parikh, D., Zitnick, C.L.: The role of features, algorithms and data in visual recognition. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2328–2335 (2010)
3. Arivazhagan, S., Shebiah, R.N., Nidhyan, S.S., Ganesan, L.: Fruit recognition using color and texture features. Journal of Emerging Trends in Computing and Information Sciences 1, 90–94 (2010)
4. Seng, W.C., Mirisaee, S.H.: A new method for fruits recognition system. In: International Conference on Electrical Engineering and Informatics, pp. 130–134 (2009)
5. Kavdır, İ., Guyer, D.E.: Evaluation of different pattern recognition techniques for apple sorting. Biosystems Engineering 99, 211–219 (2008)
6. Rocha, A., Hauagge, D.C., Wainer, J., Goldenstein, S.: Automatic Fruit and Vegetable Classification from Images. Comput. Electron. Agric. 70, 96–104 (2010)
7. Coimbra, M.T., Cunha, J.P.S.: MPEG-7 Visual Descriptors — Contributions for Automated Feature Extraction in Capsule Endoscopy. IEEE Transactions on Circuits and Systems for Video Technology 16, 628–637 (2006)
8. Lai, K., Bo, L., Ren, X., Fox, D.: A large-scale hierarchical multi-view RGB-D object dataset. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 1817–1824 (2011)
9. Sikora, T.: The MPEG-7 visual standard for content description-an overview. IEEE Transaction on Circuits and Systems for Video Technology 11, 696–702 (2001)
10. Kasutani, E., Yamada, A.: The MPEG-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval. In: International Conference on Image Processing, vol. 1, pp. 674–677 (2001)
11. Jalab, H.A.: Image retrieval system based on color layout descriptor and Gabor filters. In: 2011 IEEE Conference on Open Systems, pp. 32–36(2011)
12. Annesley, J., Orwell, J., Renno, J.-P.: Evaluation of MPEG7 color descriptors for visual surveillance retrieval. In: 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 105–112 (2005)
13. Manjunath, B.S., Ohm, J.-R., Vasudevan, V.V., Yamada, A.: Color and texture descriptors. IEEE Transactions on Circuits and Systems for Video Technology 11, 703–715 (2001)
14. Zhang, D., Islam, M., Lu, G.: A review on automatic image annotation techniques. Pattern Recognition 45, 346–362 (2012)
15. Kimura, P.A.S., Cavalcanti, J.M.B., Saraiva, P.C., Torres, R.S., Gonçalves, M.A.: Evaluating Retrieval Effectiveness of Descriptors for Searching in Large Image Databases. Journal of Information and Data Management 2, 305–320 (2011)
16. Eidenberger, H.: How good are the visual MPEG-7 features? In: SPIE & Visual Communications and Image Processing, pp. 476–488 (2003)
17. Talib, A., Mahmuddin, M., Husni, H., George, L.E.: A weighted dominant color descriptor for content-based image retrieval. Journal of Visual Communication and Image Representation 24, 345–360 (2013)
18. Kim, S.M., Park, S.J., Won, C.S.: Image Retrieval via Query-by-Layout Using MPEG-7 Visual Descriptors. ETRI Journal 29, 246–248 (2007)
19. Chang, S.-F., Sikora, T., Purl, A.: Overview of the MPEG-7 standard. IEEE Transactions on Circuits and Systems for Video Technology 11, 688–695 (2001)

20. Jiang, L., Koch, A., Scherer, S.A., Zell, A.: Multi-class fruit classification using RGB-D data for indoor robots. In: 2013 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 587–592 (2013)
21. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explor. Newsl. 11, 10–18 (2009)
22. Chang, C.-C., Lin, C.-J.: LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2, 1–27 (2011)

Content-Based Image Retrieval Using Features in Spatial and Frequency Domains

Kazuhiro Kobayashi and Qiu Chen

Kogakuin University, Faculty of Engineering,
Department of information and Communication Engineering, Tokyo, Japan
{cm14011@ns,chen@cc}.kogakuin.ac.jp

Abstract. With the rapid increase of image data due to the development of information society, the necessity of image retrieval is increasing day by day. The research result of image retrieval using Local Binary Pattern (LBP) is reported previously, which is a robust feature obtained from spatial domain. It can be considered that we can get further information as the feature of image when extracting the feature from frequency domain. In this paper, we propose a novel image retrieval algorithm to improve retrieval accuracy, which using both features obtained from spatial and frequency domains. 2-dimensional Discrete Cosine Transform (DCT) is used to calculate the feature in frequency domain. Corel database is used for the evaluation of our proposed algorithm. It is demonstrated that image retrieval using combined features can achieve a much higher search success rate compared with that of algorithms using DCT and LBP, respectively. As a result, the precision rates and recall rates of this study was higher than the preceding study. In addition, better results were obtained using the weights.

Keywords: Image Retrieval, DCT, LBP, Histogram Feature.

1 Introduction

In recent years, the spread and development of computer and digital camera, generation and collection of digital images in personal storage and WWW is increasing rapidly. In this situation, there is a growing interest in the image retrieve an image that the user is required. As a method of image retrieval, First, as a method of utilizing the text query has been previously assigned to the meta data of the query image TBIR (Text-Based Image Retrieval), Second, and the image itself query utilizing the color and texture of the image there is a method for CBIR (Content-Based Image Retrieval). TBIR is widely used, and the speed of processing is also faster. Examples of TBIR, include the image retrieval engine such as Google and is used by many users today. However, for TBIR that the text query, the burden on the people by the amount of massively growing image, such as the recent years occur. Because, since it is necessary to artificially indexed such as words and sentences to the metadata in the image in the database beforehand. Although the index, search results may differ from the ideal image because Index grant index and users demand are there may be a possibility that

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 269–277, 2015.
DOI: 10.1007/978-3-662-46742-8_25

the difference is caused by the sense of personal. Therefore, Researchers of CBIR is working hard to study towards a better search methods that Image on the Internet, video databases and other information sources.

When searching an image, first calculates the feature from the image, then comparing the feature value of the image on the query image and the database, and extracts the images by the rate of matching features. The features of the image, there is characterized derived from features and spatial domain obtained from the frequency domain. Feature of the former frequency domain, and that many methods utilizing the Gabor filter [1] used in the MPEG-7 [2], rely on them derivation algorithm of the DFT (Discrete Fourier Transform) [3]. Features obtained from spatial domain of the latter, a feature quantity based on the color and texture of the image, which is represented by the color histogram [4], color correlogram [5], descriptors such as a color moment [6]. In this study, DCT (the Discrete Cosine Transform) [7] is used for the feature of frequency domain, the features of spatial domain using a LBP (Local Binary Pattern) [8]. DCT is used converted to a standard format such as jpeg or mpeg for reasons that it has a characteristic that can increase the compression rate and the hardware implementation is easy fast algorithm. Moreover, LBP is frequently utilized by simple method which is calculated based on the texture image. Study conducted image retrieval using the LBP is done by Takala et al. [9]. In addition, there is a possibility that suggests an improvement in the success rate by using the features which are obtained from the frequency domain. The purpose of this study is to perform image retrieval using the features which are obtained from the two methods described above, to examine or consider the results.

2 Proposed Algorithm

In this paper, we propose an efficient image retrieval algorithm using combined histogram-based features which contain both spatial and frequency domain information of an image, are utilized as a very effective image feature. The flow chart of proposed algorithm is illustrated as Fig. 1. Generates pixel block in DCT, to scan the image. After converting the pixels block is quantized using a low-frequency region. DCT histogram is generated using bin obtained for each block. LBP was calculated by three operators in the same way as preceding study. LBP histograms were obtained by summing the histogram obtained from three operators. To compare the image search after registering the two features.

2.1 Discrete Cosine Transform (DCT)

DCT used in image is extended to two-dimensional DCT, represented by the transformation equation as the following eq. 1 [6].

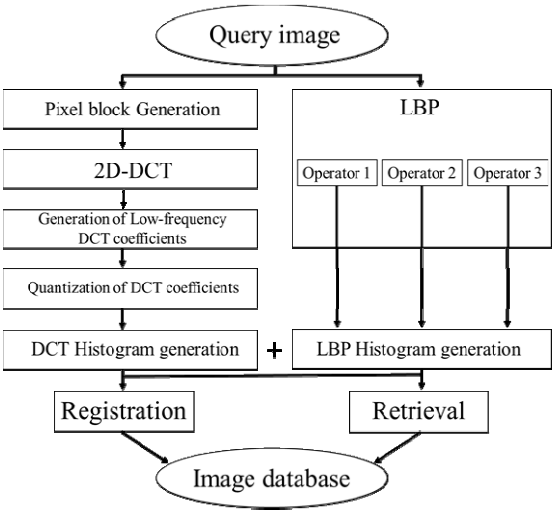


Fig. 1. Proposed algorithm

$$F(u,v) = \frac{2}{N} C(u)(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos\frac{(2x+1)u\pi}{2N} \cos\frac{(2y+1)v\pi}{2N} \quad (1)$$

$$f(u,v) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)(v) F(x,y) \cos\frac{(2x+1)u\pi}{2N} \cos\frac{(2y+1)v\pi}{2N} \quad (2)$$

Where $C(k) = \frac{1}{\sqrt{2}}(k=0), C(k) = 1(k \neq 0)$ (3)

When applying the DCT, The method is often used to perform the DCT while overlapping the small block instead of performing the DCT for the whole image data.

	0 Low	1	2	3	4	5	6	7 High
0 Low	DC	AC01	AC02	AC03	23	-6	-14	19
1	AC10	AC11	AC12	AC13	10	20	13	10
2	AC20	AC21	AC22	AC23	8	1	-20	-5
3	AC30	AC31	AC32	AC33	-9	-7	6	7
4	-9	2	5	9	-10	9	9	17
5	7	1	-10	8	4	8	1	1
6	0	0	-5	4	-7	1	-1	-7
7 High	-5	-8	8	3	2	-4	-8	0

Fig. 2. Generation of Low-Frequency DCT coefficients

Pixel block of 8×8 after conversion by the eq. 1 is represented by the DCT coefficients. After conversion, as shown in Fig. 2, the block is divided into a high-frequency component part and low frequency component parts.

We will use the features of low-frequency portion shown in Fig. 2, and we calculated number of six low frequency component. These figures are expressed as the following eq. 4 using the values of Fig. 2.

$$\left.\begin{aligned} a[0] &= AC01 \\ a[1] &= AC11 \\ a[2] &= AC10 \\ a[3] &= (AC02 + AC03 + AC12 + AC13) / 4 \\ a[4] &= (AC22 + AC23 + AC32 + AC33) / 4 \\ a[5] &= (AC20 + AC21 + AC30 + AC31) / 4 \end{aligned}\right\} \tag{4}$$

After calculating the six features of low-frequency, then performs the quantization of the DCT coefficients. When DCT coefficient is 0 or more, assigns 1 low frequency feature vectors, when the DCT coefficients is less than 0, assigns 0 to the low-frequency feature vectors. Thus calculates the feature of each block overlapping, we will continue to generate DCT histogram as in Fig. 3.

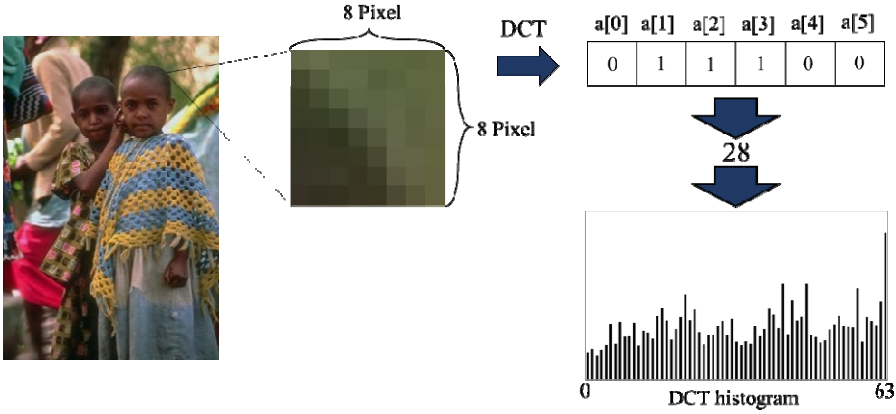


Fig. 3. DCT histogram

2.2 Local Binary Patterns (LBP)

The original LBP algorithm is proposed by Ojara et al. for local feature extraction of image [7]. The operator labels the pixels of an image by thresholding the 3×3-neighbourhood of each pixel with the center value and considering the result as a binary number. The feature of this approach is that it can be extracted local features of an image, it is hardly affected by illumination changes, it calculates the cost is low, and various applications thereof can be expected in the derived algorithm. An example of binarizing the vicinity of LBP is shown in Fig. 4.

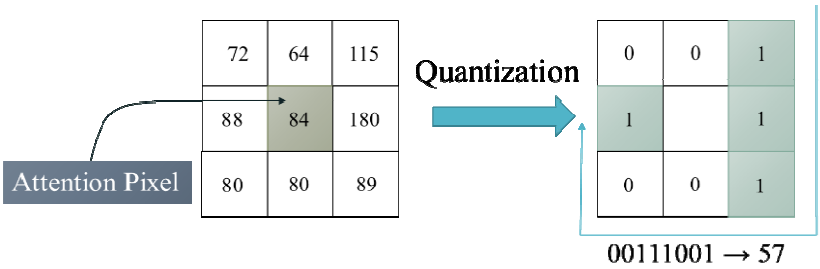


Fig. 4. The original LBP

Original LBP is possible to obtain the features of the binary number of 8 bits for LBP, to obtain the features of 0 to 255 and convert to decimal. LBP has that has been extended not limit the number of sample points instead of LBP to take neighboring pixels as described above as a sample point. Three different LBP operators as in Fig. 5 as an example the LBP of this approach is given [10].

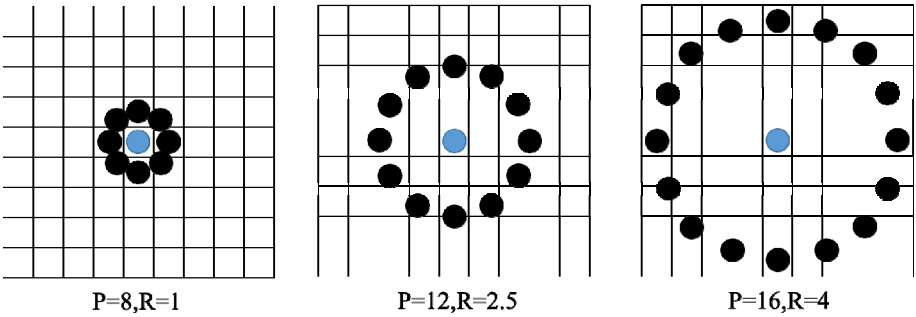


Fig. 5. Circularly symmetric neighbor sets

Sample points in order to not apply the exact pixel locations are arranged in a circle, using the bilinear interpolation in the calculation in this approach. The value taken by the LBP increases exponentially as the number of sampling points is increased. We can deal with research methods for this event that Uniform Pattern [10]. This method is intended to be reduced without damage to said frequency of feature vectors by attention 1 and 0 transition in binary value of LBP. As an example, if it takes eight interpolation points, we get the frequency of 58 instead of the frequency of 256, or, we get the frequency of 59 the rest of the pattern is stored in one frequency. An example of Uniform Pattern is shown in Fig. 6.

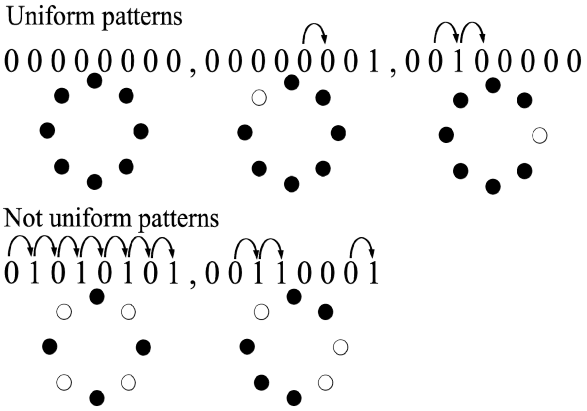


Fig. 6. Uniform Pattern

It is possible to extend the LBP by combining operators with different radii. When using N number of operators, the distance between the query and the model is expressed by the following eq. 5.

$$L_N = \sum_{n=1}^{N}(Q^n, M^n) \tag{5}$$

LBP scans the whole image like DCT, too. It is to continue to overlap the pixel block of the calculation range of LBP in the image. That the sum of the radius of different circular LBP operator respectively are generated as LBP histogram as in Fig. 7.

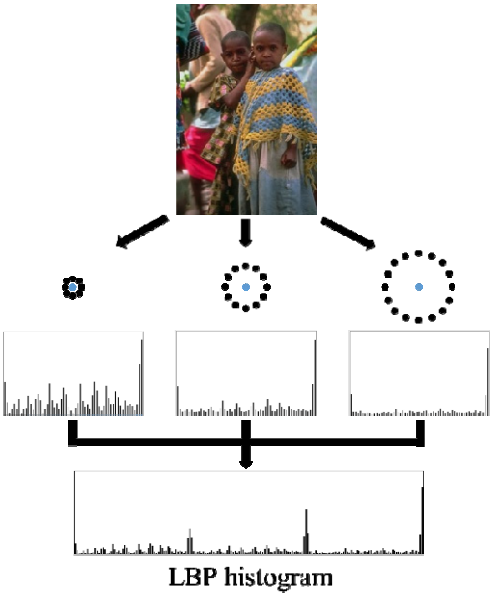


Fig. 7. LBP Histogram

3 Retrieval Metric

A distance function is needed for comparing images through their features. Most of the LBP studies got a favored a nonparametric log-likelihood statistics as suggested by Ojala et al. [10]. However, the relative L_1 measure similar to those proposed by Huang, et al. [5]. When it is compared with the likelihood with other Statistics available, relative L_1 measure is one third of the time requested by the log-likelihood. Relative L_1 measure is used this reason in Takala et al. LBP study. We also use this difference in this study. The dissimilarity measure is given in Eq. 6, where x_1 and x_2 represent the feature histograms to be compared and subscript we is the corresponding bin.

$$L_1^{relative}(x_1, x_2) = \sum_i \frac{|x_{1,i} - x_{2,i}|}{x_{1,i} + x_{2,i}} \tag{6}$$

4 Experimental Results and Discussions

4.1 Image Database

Image retrieval methods are tested on a database consisting of Wang database [11] Corel images of sizes 384×256 and 256×384. There are 10 categories of 100 images each making up 1000 images in total. Some example images are shown in Fig. 8.

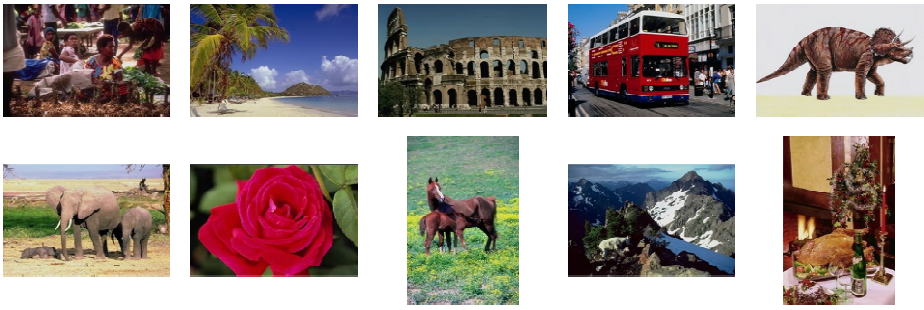


Fig. 8. Image examples from the Wang database

Africa people, buildings, flowers, elephants, mountains are chosen as a category used by an experiment. We calculate the quantity of characteristic of all the images to use in an experiment first. First, we calculate the quantity of characteristic of all the images to use in an experiment. We use Extended LBP having 3 radiuses unlike 8 interpolation point for spatial domain. It is the same as Takala et al. method, which is chosen for comparison. Frequency of each of the LBP has the Uniform pattern is reduced to 59, Extended LBP the frequency of 177 is calculated. Further, DCT features frequency domain because it has a frequency of 64, the frequency of 241 per one image is calculated one after another. We based feature calculated, to compare the feature of the query image and the database images. Enter a single query image we

want to calculate the distance compared to the feature of all of the database image. Comparison of 250 is performed per piece of the query image. Query images are made comparison of 42,500 times of an image since there 250. Retrieves the some images of the smaller distance from database image, it is confirmed whether the same category.

4.2 Experimental Results

The results of the experiments are shown in Table 1. The evaluation measures used are precision and recall. The table shows the results which using only DCT and LBP for comparison. In Takala et al. experiments, since the interpolation points are three different radii which using eight LBP operators [9], and experiments are conducted using the same LBP for comparison. Only result in the LBP, the results of the 3 operators had higher search precision and recall than 1 operator. Further compared with the result that using only DCT and LBP, the results with both methods improve the precision and recall are obtained. The addition of weight to both the feature quantity, Precision rates of 68% (recall of 13.6%) for 10images are clearly better than the other approaches.

Table 1. The evaluation results (precision/recall) for different methods

Method	10 images(%)	25 images	50images
DCT	60.4/12.1	51.3/25.7	41.2/41.2
$LBP^{u2}_{8.1}$	61.2/12.2	54.9/27.4	46.6/46.6
$LBP^{u2}_{8.1}+^{u2}_{8.2.4}+^{u2}_{8.5.4}$	64.0/12.8	56.8/28.4	47.7/47.7
DCT+$LBP^{u2}_{8.1}+^{u2}_{8.2.4}+^{u2}_{8.5.4}$	66.3/13.3	58.2/29.1	48.1/48.1
DCT+$LBP^{u2}_{8.1}+^{u2}_{8.2.4}+^{u2}_{8.5.4}$ weight	68.0/13.6	58.8/29.4	48.4/48.4

5 Conclusions

In this paper, we proposed a robust image retrieval algorithm using the features spatial domain feature frequency domain feature. Result of image retrieval accuracy by incorporating DCT to image search using a conventional LBP rises. We use both spatial domain features and frequency domain feature, which could be confirmed more that image features are obtained. However, yet we do not be derived better results for studies that blocking the image shown in Takala et al. these experiments [9]. Further good results by comparing the LBP features and DCT, wherein the image is divided into blocks can be expected. In Takala et al. experiments are also performed technique for searching a portion of an image overlapping. That approach it is necessary to perform experiments to confirm the retrieval accuracy. Also, there is a technique that has been extended to the LBP, it might be able to improve the image retrieval accuracy by incorporating them.

References

1. Manjunath, B.S., Ma, W.Y.: Texture Features for Browsing and Retrieval of Image Data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18, 837–842 (1996)
2. Manjunath, B.S., Ohm, J.-R., Vasudevan, V., Yamada, A.: Color and Texture De- scriptors. IEEE Transactions on Circuits and Systems for Video Technology 11, 703–715 (2001)
3. Sim, D.-G., Kim, H.-K., Oh, D.-H.: Translation, Rotation, and Scale Invariant Texture Descriptor for Texture-Based Image Retrieval. In: International Conferenceon Image Processing, pp. 742–745 (2000)
4. Swain, M., Ballard, D.: Color Indexing. In: Third International Conference on Computer Vision, pp. 11–32 (1990)
5. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.-J., Zabih, R.: Image Indexing Using Color Correlograms. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 762–768 (1997)
6. Stricker, M., Orengo, M.: Similarity of Color Images. In: SPIE Conference on Storageand Retrieval for Image and Video Databases, pp. 381–392 (1995)
7. Chen, Q., Kotani, K., Lee, F., Ohmi, T.: Combined Histogram-based Features of DCT Coefficients in Low-frequency Domains for Face Recognition. In: Proceeding of the Seventh International Conference on Systems and Networks Communications (ICSNC 2012), pp. 108–112 (November 2012)
8. Ojala, T., Pietikäinen, M., Hardwood, D.: A Comparative Study of Texture Measureswith Classification Based on Feature Distribution. Pattern Recognition 29, 51–59 (1996)
9. Takala, V., Ahonen, T., Pietikäinen, M.: Block-Based Methods for Image Retrieval Using Local Binary Patterns. In: Kalviainen, H., Parkkinen, J., Kaarna, A. (eds.) SCIA 2005. LNCS, vol. 3540, pp. 882–891. Springer, Heidelberg (2005)
10. Ojala, T., Pietikainen, M., Mäenpää, T.: Multiresolution Gray-Scale and RotationInvariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971–987 (2002)
11. Wang Group: Modeling Objects, Concepts, Aesthetics, and Emotions in Images, `http://wang.ist.psu.edu/docs/home.shtml`

Feature Extraction for Java Character Recognition

Rudy Adipranata, Liliana, Meiliana Indrawijaya, and Gregorius Satia Budhi

Informatics Department, Petra Christian University
Siwalankerto 121-131, Surabaya, Indonesia
{rudya,lilian,greg}@petra.ac.id, meiliindrawijaya@gmail.com

Abstract. Feature extraction is very important in the process of character recognition. A good feature of the character will increase the level of accuracy for the character recognition. In this research, the feature extraction experiment of Java characters is done, where those features could be used for Java character recognition later. Before performing the process of feature extraction, segmentation is performed to get each Java character in an image, and followed by skeletonization process. After skeletonization process, feature extraction is done including simple closed curve, straight lines and curve. Several experiments was done using various parameters and Java characters in order to obtain the optimal parameters. The experiment results of simple closed curve and straight line feature extraction are quite good, respectively reached 82.85% and 74.28%. However, the result of the curve detection is still not good, only reached 51.42%

Keywords: feature extraction, Java character, skeletonization.

1 Introduction

One of the ethnic groups found in Indonesia is the Javanese with a culture that is known as the Javanese culture, which has many aspects. One is the use of Java characters on writing. To preserve the culture in the form of Java character, conservation efforts need to be done so that the culture is not extinct, one is to digitize documents bearing Java characters especially old documents that have high historical values. Java characters consist of basic characters, numbers, complementary characters and others. In this paper, we conduct research on feature extraction of Java characters. The features obtained later will be used for the Java character recognition in order to digitize the documents. These experiments of feature extractions need to be done in order to find different features for each Java character, because some of the Java characters have similar form to each other. Input to the system is an image that contains handwritten Java characters. Image segmentation is performed to separate the existing characters, and then we do the skeletonizing before feature extraction. The feature extraction was conducted on the number of lines, simple close curves and open curves.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 278–288, 2015.
DOI: 10.1007/978-3-662-46742-8_26

2 Skeletonizing

Skeletonizing is one of image processing that is used to reduce the pixels of an image while maintaining information, characteristics and important pixels of the object. This is implemented by changing the initial image in binary into skeletal representation of the image. The purpose of skeletonizing is to make simpler image so that the image can be analyzed further in terms of shape and suitability for comparison with other images. Problems encountered while doing skeletonizing is how to determine the redundant pixels and to maintain important pixel. This process is more similar to erosion process where erosion can lead to an unexpected region deleted, which is not expected in the process of thinning. Skeleton must remain intact and have some basic properties such as [1]:

- Must be composed of several thin regions, with a width of 1 pixel.
- Pixels that form the skeleton should be near the middle of the cross section area of the region.
- Skeletal pixel must be connected to each other to form several regions which are equal in number to the original image.

There are two requirements that are used to determine whether a pixel can be removed or not. The first requirement is as follows [1]:

- A pixel can be removed only if it has more than one and less than 7 neighbors.
- A pixel can be removed only if it has the same one connectivity.
- A pixel can be removed only if at least one of the neighbors who are in the direction of 1, 3, or 5 is a background pixel.
- A pixel can be removed only if one of the neighbors who are in the direction of 3, 5, or 7 is a background pixel.

The second requirement is similar to the first but differ in the last two steps:

- A pixel can be removed only if at least one of the neighbors who are in the direction of 7, 1, or 3 is a background pixel.
- A pixel can be removed only if one of the neighbors who are in the direction of 1, 5, or 7 is a background pixel.

3 Simple Closed Curve

Simple closed curve is a simple curve that both endpoints coincide [2]. Examples of simple closed curve can be seen in Figure 1.

Fig. 1. Simple close curve

To detect a closed curve can be done by using the flood fill algorithm. Flood fill algorithm has three main parameters, namely the start node, the target color and color replacement. This algorithm searches all nodes in the array are connected to the start node through the path of the target color and then replace it with a replacement color. Flood fill based algorithm by using recursion can be written as follows:

1. If the node is not the same color with the color of the target, the return
2. Set the color of nodes into a replacement color.
3. Run Flood-fill (one step to the west of the node).
   Run Flood-fill (one step to the east of the node).
   Run Flood-fill (one step to the north of the node).
   Run Flood-fill (one step to the south of the node).
4. Return.

4 Hough Transform

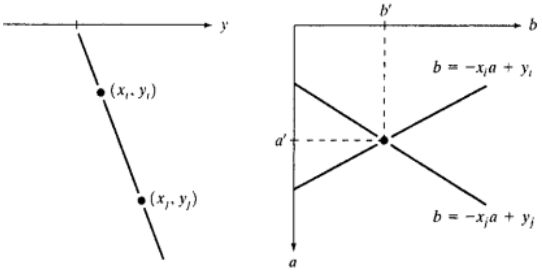
Hough Transform is a technique for determining the location of a shape in the image. Hough Transform was first proposed by P.V.C Hough [3], and implemented to detect the lines in the image by Duda and Hart [4].

Hough Transform maps the points in the image into the parameter space (Hough Transform space) based on a function that defines the shape that wants to be detected, and then takes a vote on an array element called the accumulator array. Hough Transform is generally used to perform the extraction of lines, circles or ellipses in the image, but in its development, Hough Transform can also be used to the extraction of more complex shapes. Hough Transform is used to detect the straight lines that satisfy the Equation 1 and 2:

$$y = a\,x + b \tag{1}$$

$$b = -x_1\,a + y_1 \tag{2}$$

By changing the Equation 1 to 2, each edge point (x, y) on an image will result in single line equation parameters (a, b). The points on the same line will have the value of the parameter that cross at a point (a, b) in the parameter space as shown in Figure 2.



(a) (b)

Fig. 2. (a) xy Area (b) Parameter space

At first, the value of the accumulator is initialized with zero. For each point (x, y) which is edge of the object in the image, the value of b will be calculated according to the Equation 2 and will be rounded to the nearest value that is allowed in the accumulator. Accumulator value will be incremented for each point that satisfies the equation with the values of a and b given in accordance with the Equation 3:

$$A(a,b) = A(a,b) + 1 \tag{3}$$

Each edge point has line parameter mapped in the accumulator. The higher the value in the accumulator, the greater the likelihood of a line is detected in the image. Duda and Hart proposed a polar equation for a line with a parameter ρ and orientation θ [4] (Equation 4). Illustration can be seen in Figure 3.

$$\rho = x \cos\theta + y \sin\theta \tag{4}$$

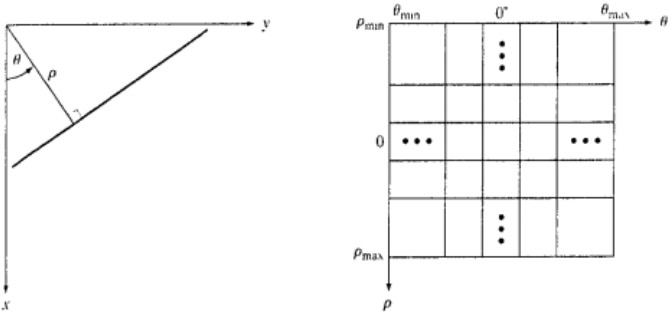


Fig. 3. (a) Normal representation of a line. (b) Parameter space (ρ, θ)

In the same way with the standard Hough transform, each point in the image is mapped into the accumulator for each value ρ and θ which satisfy the Equation 5.

$$A(\rho,\theta) = A(\rho,\theta) + 1 \tag{5}$$

The range of values for the angle θ is ± 90 as measured by the x-axis. While the range of values of ρ is $\pm\sqrt{2}\,D$, where D is the distance between the vertex on the image [5].

In general, the Hough Transform method consists of three basic steps:

- Each pixel of an image is transformed into a curve parameter of the parameter space.
- Accumulator with cell arrays placed on the parameter space and each pixel images provide a value to the cells in the transformation curve.
- Pixel with a local maximum value is selected, and the coordinates of the parameters used to represent a segment of the curve in image space.

4.1 Parabolic Curve Detection Using Hough Transform

In the image of the actual object, the curve can be in any orientation. Parabolic curve with rotation can be detected by using an algorithm based on coordinate transformation of parabolic equations. In standard parabolic curve detection, there are four parameters involved, namely the point (x_0, y_0), orientation (θ), and the coefficient which contains information about the parabolic curvature. However, Jafri and Deravi proposed an algorithm to detect parabolic curve in any orientation using only three parameters [6]. The parameters are the point (x_0, y_0) and orientation θ. Using this algorithm, all parabolic curves in various positions can be detected by using 3D accumulator. This approach uses a point on the curve as a parameter which also shows the position of maximum curvature of the parabolic curve. Sobel operator is used for the gradient approach. To detect the parabolic curve in any orientation, a coordinate transformation matrix is used to derive a new parabolic equations involving parabolic curve orientation.

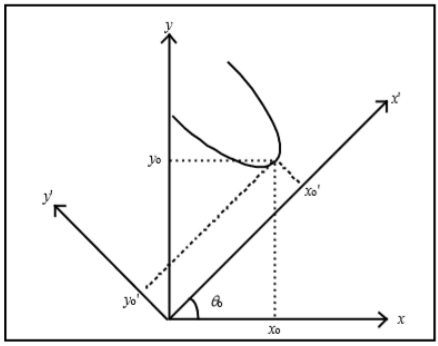


Fig. 4. Parabolic curve

Parabolic curve with a specific orientation angle is shown in Figure 4. (x', y') coordinates is the (x, y) coordinates rotation by θ degrees with the center coordinate system as the axis of rotation. The vertex of parabola is (x_0', y_0') at the (x', y') coordinates or (x_0, y_0) in the (x, y) coordinates. The equation of the parabola in the (x', y') coordinates can be written in Equation 6 [6].

$$(y' - y_o') = p\,(x' - x_o')^2 \tag{6}$$

Standard two dimensional geometry matrix for counter-clockwise rotation with θ angle transformation is shown in Equation 7.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{7}$$

By substituting the value of x, y, x_0 and y_0 in Equation 7 to Equation 6, the parabolic Equation 6 can be written as:

$$\begin{aligned} &(-x\sin\theta + y\cos\theta) - (-x_o\sin\theta + y_o\cos\theta) \\ &= p\,[(x\cos\theta + y\sin\theta) - (x_o\cos\theta + y_o\sin\theta)]^2 \end{aligned} \tag{8}$$

And the value of differentiation of this equation is:

$$-\sin\theta + \frac{dy}{dx}\cos\theta$$

$$= 2p\,[(x\cos\theta + y\sin\theta) - (x_o\cos\theta + y_o\sin\theta)]\,.\left[\cos\theta + \frac{dy}{dx}\sin\theta\right] \quad (9)$$

By substituting Equation 9 into 8, a new relation to the parabola vertex and the orientation (x_0, y_0, θ) is [6]:

$$y_o = \left[\frac{k_1(x\cos\theta + y\sin\theta) + (x\sin\theta - y\cos\theta)}{k_1(\sin\theta - \cos\theta)}\right]$$

$$\frac{(k_1\cos\theta + \sin\theta)}{(k_1\sin\theta - \cos\theta)}x_o \quad (10)$$

where k_1 is

$$k_1 = \frac{-\sin\theta + \frac{dy}{dx}\cos\theta}{2(\cos\theta + \frac{dy}{dx}\sin\theta)}x_o \quad (11)$$

From the above relationship, it can be seen that the use of three dimensional accumulator arrays is sufficient to detect a parabola in various orientations.

5 Experimental Results

Several types of experiments has been performed, such as simple close curve detection, line detection and curve detection on several sample scanned document images

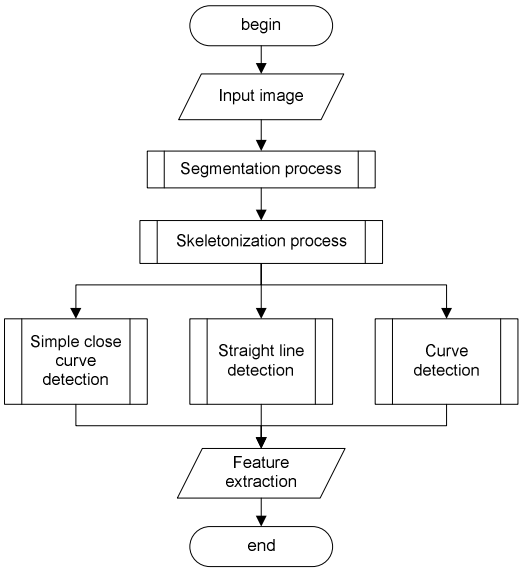
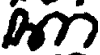
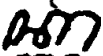


Fig. 5. Flowchart of the system

of Java characters. The overall process can be seen in Figure 5. At first, system will accept input images that contain Java characters. Then segmentation process will be done to separate each character. For each character, skeletonization process will be done before detection process.

The following are the results of the simple close curve detection. The number of simple close curve contained in several sample images can be seen in Table 1.

Table 1. Experiments results of simple close curve detection

No	Java Character	Number of simple close curve detected
1		0
2		2
3		3
4		3
5		0
6		2
7		1
8		0
9		0
10		4
11		0
12		2
13		1
14		1
15		2
16		1
17		1
18		2
19		0
20		0
21		2
22		0
23		1
24		1
25		1
27		1

From the experiment results in Table 1, there are incorrect calculations of the number of close curve, that is the sample number 2, 3, 4, 6, 10, 12, 15, 16, 18, 21, 24, and 25. This error rate reaches 44.44%. Some existing problems are:

1. Two parallel lines are too close together, so it becomes a closed curve after the thinning process.
2. There are areas that are too close together, so it becomes a closed curve after the thinning process.
3. The presence of noise in the form of a closed curve with a small size.

To solve the problem number three, we add minimum area parameter which used to limit the value of the minimum area of a closed area that can be considered as the actual close curve, not just as a noise. To overcome the problem of numbers 1 and 2, we divide the image into two equal segments (top and bottom). By using minimum area parameter and dividing image, we got better experiment result. Closed curve is considered as noise if the area less than or equal to 2 pixels.

The following is experiment of line detection by using several parameters. The parameters are:

- Theta resolution: Resolution angle in degrees.
- Threshold: The minimum value limit value in the accumulator array can be expressed as a line. Threshold consists of two types, namely by percentage and by value.
- Min line length: The value of the minimum length of a line segment can be considered to be a line.
- Max line gap: The minimum distance between the lines can be detected as a line.

From the experiment results in Table 2, we can conclude that the smaller the value of theta resolution, more lines can be detected, and also we can see that theta resolution value of 1 gives the best results.

We also do experiments on changing of threshold parameter. From experiment results, it can be concluded that if the threshold is too small, will causes noise be detected. However, if the threshold is too large causing a short line will not be detected. The threshold value for 7 pixels provide relatively good results and stable for line detection.

From other experiment on changing of line length, we get results that the optimum line length ranges between 4 - 5 pixels.

From experiments on changing maximum line gap, it can be concluded that if the max line gap is too small will causes many lines are detected. And from the results, maximum line gap of 4 pixels will give good results.

The following is experiment of curve detection by using several parameter values. The experiment result can be seen in Table 3.

Table 2. Experiment results of line detection

No	Parameter	Image	Thinned image	Result image	Union of line detected and thinned image	Number of line detected
1	Theta resolution: 1 Neighbor: 4 Threshold: 5 Min line length: 4 Max line gap: 4					4
						5
2	Theta resolution: 2 Neighbor: 4 Threshold: 5 Min line length: 4 Max line gap: 4					8
						8
3	Theta resolution: 3 Neighbor: 4 Threshold: 5 Min line length: 4 Max line gap: 4					7
						7
4	Theta resolution: 5 Neighbor: 4 Threshold: 5 Min line length: 4 Max line gap: 4					2
						5
5	Theta resolution: 7 Neighbor: 4 Threshold: 5 Min line length: 4 Max line gap: 4					3
						3
6	Theta resolution: 1 Neighbor: 4 Threshold: 5 Min line length: 4 Max line gap: 4					5
						2
7	Theta resolution: 3 Neighbor: 4 Threshold: 5 Min line length: 4 Max line gap: 4					4
						3
8	Theta resolution: 5 Neighbor: 4 Threshold: 5 Min line length: 4 Max line gap: 4					4
						2

Table 3. Experiment results of curve detection

No	Parameter	Image	Result Image	Number of curve detected
1	Theta Step: 4 Neighbor: 7 Threshold: 9		 	4 0
2	Theta Step: 5 Neighbor: 7 Threshold: 9		 	3 0
3	Theta Step: 6 Neighbor: 7 Threshold: 9		 	4 0

The parameters tested in the detection curve are theta step, neighbor and threshold value. Theta step value affects the increment value of angle (theta), the line will be detected every n-degree value of angle. From the experiment results, it can be concluded that the smaller value of theta step, the detail lines can be detected, and the theta step value 1 gives the best results.

Neighbor parameters affect the width of the window for detecting local maxima. From experiment results, the smaller the value of the neighbor, more lines can be detected because of the larger peak, and the optimum value of neighbor is 7, it will get the best result.

The threshold parameter affects the number of crossovers in the parameter space. From experiment results, it can be concluded that if the threshold is too small, will cause noise is detected. However, if the threshold is too large causing a short line is not detected. The optimum threshold value of 7 pixels gives relatively good results.

By using the parameters obtained from each experiment, we conduct experiment on a number of Java characters document images to get the number of simple closed curve, lines and curves as well as a comparison with the calculation of the number manually. The results of the simple closed curve detection reaches 82.85%, and the result of line detection reached 74.28%. While the result of curve detection is still low, reaching only 51.42%.

6 Conclusion

In this paper we have conducted experiment on feature extraction of Java character document images that later will be used for the detection of Java character. Based on the experiment results, it can be concluded that the optimum parameters for the

detection of simple closed curve is a minimum area of 2 pixels, while the optimum parameters for detecting the line is the theta resolution value of 1 pixel, threshold 7 pixels, line length ranges between 4 - 5 pixels, and maximum line gap 4 pixels. For detection of curves, the best parameters are theta step 1, neighbor 7 pixels and the threshold 7 pixels. From experiment on the samples of Java character document images, the results of simple closed curve and line feature extraction are quite good, respectively reached 82.85% and 74.28%. However, the result of the curve detection is still not good, only reached 51.42%.

Acknowledgment. We thank to PT Coordination of Private Higher Education Region VII, East Java, Indonesia for funding this research by Research Competitive Grant DIPA-PT Coordination of Private Higher Education Region VII, East Java, Indonesia fiscal year 2015 entitled (in Bahasa Indonesia) "Aplikasi Pengenalan Aksara Jawa Guna Digitalisasi Dokumen Beraksara Jawa Untuk Mendukung Pelestarian Budaya Nasional".

References

1. Parker, J.R.: Algorithm for image processing and computer vision. John Wiley & Sons, Inc., New York (1997)
2. Adjie, N., Maulana.:Mathematics Problem Solving. UPI Press, Bandung (2006)
3. Hough, P.V.C.: Machine Analysis of Bubble Chamber Pictures. In: Proc. Int. Conf. High Energy Accelerators and Instrumentation (1959)
4. Duda, R.O., Hart, P.E.: Use of the Hough Transformation to Detect Lines and Curves in Pictures. Comm. ACM 15, 11–15 (1972)
5. Gonzalez, R., Woods, R.: Digital Image Processing, 3rd edn. Prentice Hall, New Jersey (2008)
6. Jafri, M., Deravi, F.: Efficient algorithm for the detection of parabolic curves. In: Vision Geometry III, pp. 53–62 (1994)

Fast Performance Indonesian Automated License Plate Recognition Algorithm Using Interconnected Image Segmentation

Samuel Mahatmaputra Tedjojuwono

Bina Nusantara University, School of Information Systems, Jakarta, Indonesia
smahatmaputra@binus.edu

Abstract. A reliable and fast algorithm is needed to implement a License Plate Recognition in a real life vehicle traffic environment. This paper proposed a hybrid algorithm of interconnected image segmentation and contour extraction to locate and identify license plate for a given image. A pre-processed image will be segmented in several small images based on contour detection algorithm. The location of each segmented image's centre was then plotted into a Euclidean plane to get their relative positions. The distance and position of the segments evaluated in search of interconnected segments. This set of segments then will be fed into a cost function that determines its level of relevancy as a plate number. Set with highest score went through character recognition process to extract the meaning. Proposed work was tested in Indonesian license plate system with a high rate of correctness and average processing speed of 100 millisecond.

Keywords: License Plate Recognition, Contour Detection, Interconnected-Segments.

1 Introduction

The use of Automatic License Plate Recognition (ALPR) in real live environment is very much increasing in need. This is supported by a study in its usage for the security, surveillance and law enforcement system [1]. The study showed that although it is considerably expensive, ALPR may increase the traffic surveillance system productivity. Moreover, to perform in real live situation the system should incorporate two major characteristics which are fast and precision. These characteristics are very much the centre of discussion in all works and researches in improving ALPR system. This paper will focus on the first part of that characteristic without neglecting its accuracy level as well. An agile and fast performance recognition system is proposed using interconnected image segmentation of image features extraction technique. The system was tested in particular for major Indonesian license plate system whereas it is a one-line horizontal plate with white digit color and black background. The result of proposed algorithm showed promising outcomes in term of processing time (efficiency) with reasonable detection accuracy. Apart from its fast and precision performance,

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 289–300, 2015.
DOI: 10.1007/978-3-662-46742-8_27

the proposed system needs to work under certain reasonable condition such as light and camera position with properly applied parameters' values.

2 Previous Work

It was concluded from a recent study that each methods in solving ALPR system will have their own pros and cons [2]. This study stated that using pixel values with template matching will be simple and straight forward, however it is prone to process not important pixel and noise. On the other hand using feature extraction various methods will be robust and considerably fast, however it should dealt with some of the non robust features that could degraded accuracy performance of the algorithm. Some previous experiments on template matching techniques used fuzzy logic and neural network aims to identify number in the plate [3], with a promising accuracy rate sitting at 93.7%. However as quoted from their experimental result in part five, the localization process used 0.4 sec (400 millisecond) on top of its character recognition process which took around 2 sec (2000 millisecond), thus in total of around 2400 millisecond (2.4 second) per plate number recognition. This will not be an ideal approach to be applied in live traffic system.

Another work that was conducted under Indonesian license plate system is [4] which uses ANN (artificial neural network) in supervised system with a regular training time of 10 minutes using 400 input units. This system required human supervision and huge number of input and training set in its plate localization process. Another novel work [5] uses sliding concentric windows to locate the license plate and a two layer PNN (probabilistic neural network) for optical characters recognition. This approach reached a very efficient time of 276 millisecond (0.28 second) in average performance of 85% accuracy, however it had never been tested using Indonesian license plate.

A real time ALPR system was proposed [6] for Indonesian plate system using Fourier transform and HMM (Hidden Markov model). Its performance was averagely around 2844 millisecond (2.8 second) for a complete plate recognition process with success rate of 84.38%. As character recognition methods, some of previous work used template matching for already segmented and not tilted character in combination of other localization technique[7] [8], the latest used a better feature extraction using Euler number grouping to locate the license plate in the image. It was called the hybrid fuzzy [8] in which was tested under Malaysian license plate system that has similarity with the proposed experimental environment in Indonesia. Its performance are 95.5% for blurred or distorted image and 100% for ideal image, with total processing time at maximum of 2.45 seconds – which is considerably slow to be implemented at live traffic.

3 Proposed Methodology

The proposed algorithm is divided into three main parts: (i) Image preprocessing (ii) License plate localization (iii) License plate character recognition. The detail block

diagram can be seen at Figure 1. To achieve both reliable and fast ALPR system performance, the system used a preprocessed image as an input. Embedded to the system is an image preprocessing modules that converts raw color image to a gray level image. A smoothing processed then applied to the image using Gaussian technique in order to minimize noise.

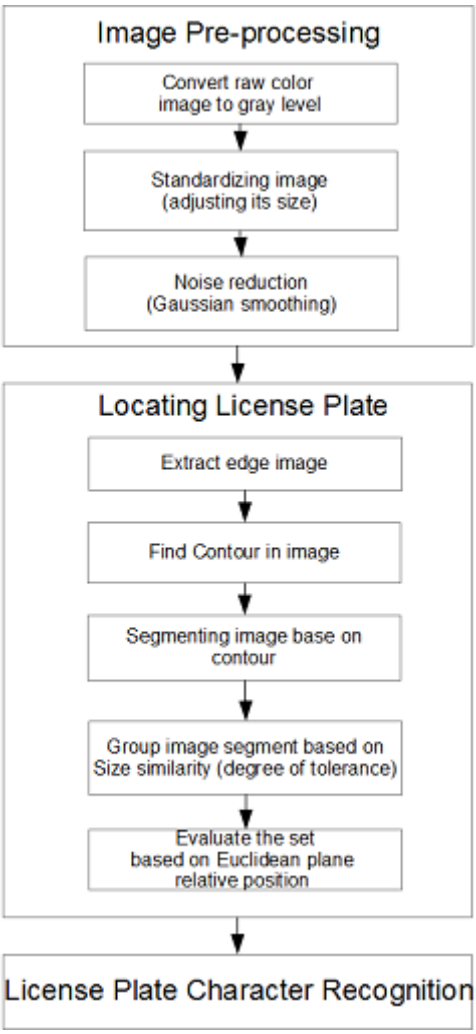


Fig. 1. Main algorithm of the proposed work

The preprocessed image will then fed into the license plate localization process. This process will segments the image I into several small images I_i based on the findings of contours c_i. However, to extract contours c_i from the image, edge detection was first applied using *canny* edge detection technique. A certain ratio evaluation of c_i width and height $r(w:h)_i$ is needed to determine whether the I_i is a possible license

plate number or not. I_i with certain degree of r_i similarity will then are grouped together in one set S_j.

The algorithm of interconnected segmentation forming the set S_j is described by the following pseudo-code:

```
function interconSegment (image[ ] Segments) {
  set image iSegment[ ]
  set image neighbors[ ]
  set decimal tolerateDeviation
  for each image in Segments {
    for each image in Neighbors {
      If Segments[i].height == neighbors[j].height
       & Segments[i].width == neighbors[j].width
& Segments[i].height/Segments[i].width==tolerateDeviation
       {
         add Image to iSegment;
       }
    }
  }
  return iSegment[]
}
```

The code compare each image in the *Segments* collection to all its closest array of *neighbors*' image segment. It also incorporate toleration degree *tolerateDeviation* in the process of comparing *width* and *height* before adding the similar size of segment to a certain *iSegment* collection.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i A_i}{\sum_{i=1}^{n} A_i} \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i A_i}{\sum_{i=1}^{n} A_i} \tag{1}$$

The grouping process started with finding *centroid* or the center point of an image segment using formula described in (1) whereas x and y axis will be acquired through calculating total moments in x and y divided by its axis total area respectively. The *centroid* position C_i then evaluated using its relative two dimensional coordinate $p(x, y)_i$ to the nearest I_i neighbors in the *I Euclidean* plane. The algorithm then evaluate S_j number of member. The largest set will then be considered as possible license plate. The location of S_j is then determined by the position of its left most member's $p(x, y)_i$. This technique then proposed to be called interconnected image segmentation. Each I_i in the selected S_j will then translated into ASCII character using a static character classifier technique proposed by [9].

4 Result

4.1 Experimental Result and Parameters

The experimental work used 100 samples of major Indonesian vehicle license plate (black background with white Latin font). Images were taken randomly at both indoor and outdoor environment. It should have reasonable amount of lighting and the camera positioned as such be perpendicularly enough with the license plate two dimensional planes. Figure 2 to Figure 6 describe the result of two main parts of the proposed work: (i) Image pre-processing (ii) License plate localization. Figure 2 shows the original image containing license plate. This image then will be converted into gray image for efficiency in finding edges.

Fig. 2. Original image as an input

The grey level image was then applied smoothing technique to reduce noise. This smoothed gray image then went through edge extraction algorithm. From the experiment, the size of searching field is minimized using this technique which in return minimizes the processing time. Figure 3 shows the result of applying edge extraction to the gray level image. The proposed work relying on the edge detected around each object inside the picture, including the letters of the plate number that will form a continuous line which is called 'contour'. The edge image then fed into contour detection function to mark any blob position resulting from detecting interconnected edges' pixels. All the contour/blob position will be recorded inside an array list to be evaluated.

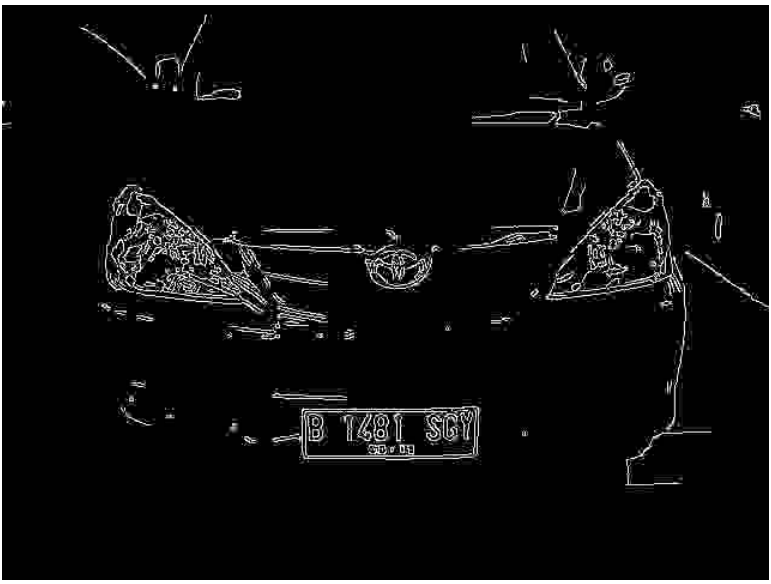


Fig. 3. Edge extracted image

Fig. 4. Segments candidates on original image (marked with orange squares)

The results of all detected contours were then distributed into several image segments. These image segments as the candidates of the plate letters are shown in Figure 4 marked by orange rectangles overlaid on top of the original image. Thus the algorithm cut through the process of searching plate number like rectangular shapes.

Fig. 5. Segment candidate location in the two dimensional plane

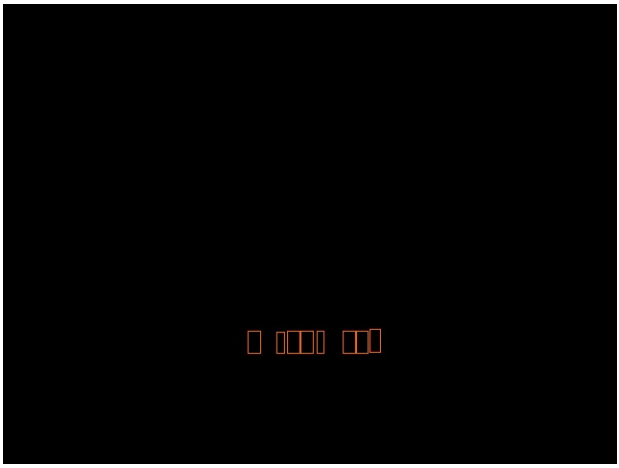


Fig. 6. License plate characters' locations are well identified

Figure 5 shows all the segments' locations in the two dimensional plane without the overlaid image. This image showed a scattered position of candidates, some of which are inside the others which will be eliminated during the filtering process. To be selected as a plate number letters, the image segment candidates should satisfy certain letter like ratio of width and height, along with their relative position to their neighboring candidates. The filtered result of all image segments is shown in Figure 6 where all the license plate characters' locations are well identified. This result of license plate localization, that is a set of image segments taken from the license plate character images, will be fed in to a character recognition function. The function will translates image into ASCII characters. Figure 7, 8, 9 and Figure 10 show the results of successfully localized and recognized license plates in average of 100 millisecond for image size of 480x640 pixels (recommended size in pixel).

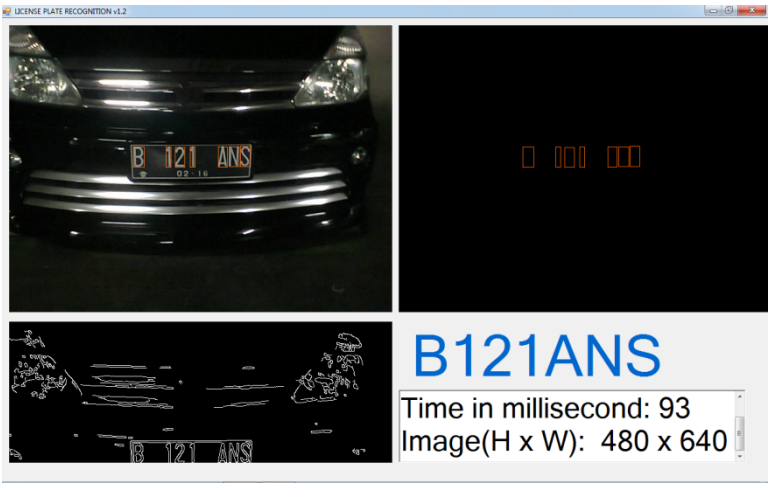


Fig. 7. Application of plate reading at outdoor-night

From figure 7 we can see the prototype of ALPR System application. Top left and right image windows are the result of plate localization with and without the original image overlaid. At the bottom right it shows the result of characters recognition process along with time in millisecond and the image dimensional in pixel unit. At the bottom left image window is the edge extraction result.

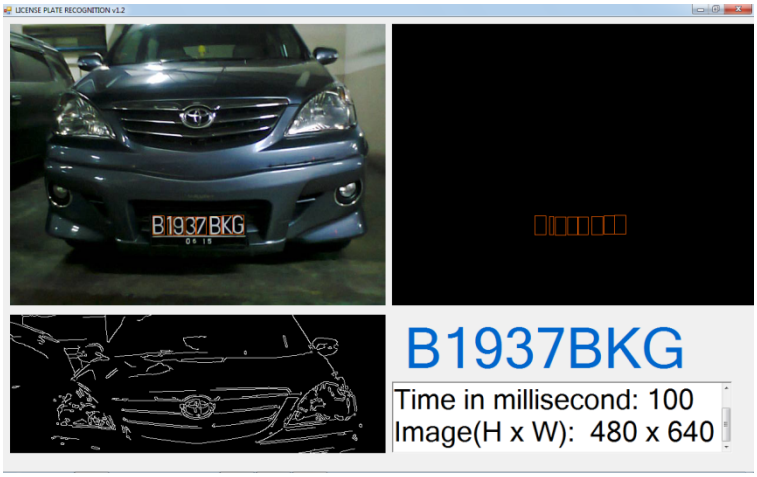


Fig. 8. Application of plate reading indoor-1

Both Figure 8 and Figure 9 show the Application of the ALPR system at indoor environment. Moreover, both showed consistent result of accuracy and performance of around 100 millisecond. Figure 10 on the other hand is an example of running the ALPR application at outdoor environment with enough sun light. However, the result

shows yet again a consistent performance both on accuracy and -more importantly on a live implementation – on processing speed which is just below 100 millisecond.

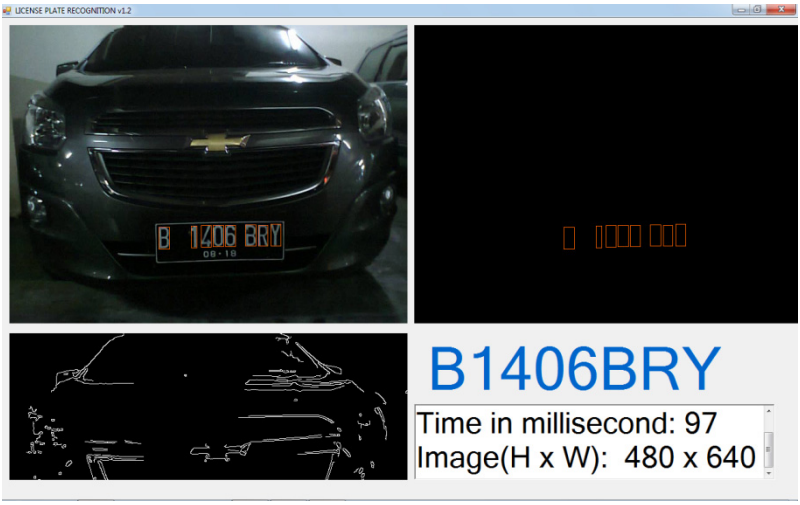


Fig. 9. Application of plate reading indoor-2

Fig. 10. Application of plate reading at outdoor bright sun light

The parameter values that are used in the experiment can be found inside Table 1. These parameters used in the smoothing process, contour extraction function, image segments filtering and interrelating images process:

Table 1. Parameters value of the proposed work

Parameters	Suggested values
Kernel size (smoothing)	3
Approximated poly	0.001
Valid character dimension ratio R	$2.0 \leq R \geq 3.7$
Tolerated deviation (2D Euclidean plane position)	0.19

Moreover, the ALPR Application and the hardware specification were built and tested under the following specifications:

- Processor: Intel Core i5 @ 2.40 GHz
- RAM: 4 GB (2.16 GB usable)
- Intel(R) HD Graphics Family Adapter RAM 916.37 MB (960,880,640 bytes)
- Display resolution 1280 x 800 x 58 hertz
- Image camera 2MP
- Windows 7 OS 32-bit
- Microsoft .Net Framework 4.5.x
- Microsoft Visual Studio C# 2010
- OpenCV library

The performance of the proposed algorithm is promising both in speed and its accuracy of locating the plate candidate. Randomly taken image were used as a test case, whereas all these samples taken from indoor and outdoor lighting condition with 97% of localization accuracy rate and 85% of character recognition rate. The average time elapsed are 95 millisecond over a 480x640 image dimension in pixel.

4.2 Problem Encountered

Figure 11 is an example of problem that was encountered during the experiment. In this test case the supposedly translated “B1629NFK” plate was missed translated as “B\629NFK”. The slightly tilted number “1” - as shown in Figure 12 - was translated as a back slash “\” sign. In this experiment this problem was referred as the tilted image problem. This problem might occur when the image taken at angle not perpendicularly close to the plate plane. Another problem that could be encountered is referred as the ornament problem. This problem might impact sample with ornaments (front grill, other printed materials, etc) that is present in the image along with the plate sample. Figure 11 shows an example of this particular problem. The grill of a minibus in the image Figure 13 has a form that is significantly satisfied a license plate criteria. This can be seen clearly from the edge image extraction result in Figure 14.

Fig. 11. Tilted image problem

Fig. 12. Tilted image problem detail

The contours generated from the edge of the bus front grill is in the correct letter (w x h) ratio and all its segments lied in on almost straight horizontal line (points by yellow arrow) with segments number that is significantly higher than the true license plate. Thus, the contours generated from the grill has miss led the algorithm to translated it as ROI (region of interest). This in result has caused the algorithm miss located the license plate position completely Figure 13 whereas every rectangular with letter like size has been incorrectly recognized as letter image segments of a license plate.

Fig. 13. The ornament problem

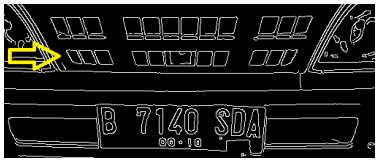


Fig. 14. The ornament problem – edge extraction

5 Conclusion

The proposed work presents a fast performance license plate localization that was tested on 100 samples of Indonesian license plate system taken from various indoor and outdoor conditions both day and night. The result of the proposed algorithms was promising as the average time for a complete ALPR process of image pre-processing, license plate localization, and character recognition took 100 millisecond in average. The accuracy of localization process was 97% whereas the character recognition came up with 85% accuracy.

Moreover, in the future the work still needs to be equipped with transformation process as to deal with tilted license plate images as a result of a poor camera positioning. The algorithms also will incorporate filtering techniques to suppress the ornament problem previously described. Methods and algorithms will be designed and implemented carefully as to maintain the fast performance of the proposed system. Future work would also compare – perceived to be the most time consuming – process of license plate character recognition with the other part of ALPR system: (i) Image preprocessing (ii) License plate localization.

References

1. Manson, T.: Automatic License Plate Recognition Systems. Law and Order 54(12), 72–77 (2006)
2. Du, S., Ibrahim, M., Shehata, M., Badawy, W.: Automatic License Plate Recognition (ALPR): A State-of-the-Art Review. IEEE Transactions on Circuits and Systems for Video Technology 23(2) (2013)
3. Chang, S.L., Chen, L.S., Chung, Y.C., Chen, S.W.: Automatic License Plate Recognition. IEEE Transaction on Intelligent Transport System 5(1) (2004)
4. Arya, F., Suwardi, I.S.: License Plate Recognition System for Indonesian Vehicles. E-Journal Teknologi Industri. B-87 (2012)
5. Anagnostopoulos, C.N.E., Anagnostopoulos, I.E., Loumos, V., Kayafas, E.: A License Plate-Recognition Algorithm for Intelligent Transportation System Applications. IEEE Transaction on Intelligent Transport System 7(3) (2006)
6. Hermawati, F.A., Koesdijarto, R.: A Real-Time License Plate Detection System For Parking Access. Telkomnika: Indonesian Journal of Electrical Engineering 8(2), 97–106 (2010)
7. Rahman, C.A., Badawy, W., Radmanesh, A.: A Real Time Vehicle's License Plate Recognition System. Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS 2003 (2003)
8. Faqheri, W.A., Mashohor, S.: Real-Time Malaysian Automatic License Plate Recognition (M-ALPR) using Hybrid Fuzzy. IJCSNS International Journal of Computer Science and Network Security 9(2) (2009)
9. Smith, R.W.: History of the Tesseract OCR engine: what worked and what didn't. In: Proceeding SPIE 8658. Document Recognition and Retrieval XX., vol. 865802 (2013)

Image Recognition

A Study of Laundry Tidiness: Socks Pairing Using Video and 3D Sensors

Maiya Kazuki*, Miyoshi Tsutomu, and Hirose Daiki

Graduate School of Science and Technology, Ryukoku University, Shiga, Japan
mijosxi@rins.ryukoku.ac.jp

Abstract. Recent years, house-robots have become popular with the advances in robot technology, but there is no alternative for the tidiness of laundry by robots. We analyzed the functions required to tidy up laundry by robots, and thought that the pairing of socks is important. The purpose of this study is to develop a system for pairing socks automatically by robots. This paper provides the recognition part of this system that includes deforming detection and the similarity determination of socks by the use of image processing and 3D information. The results show that the system could detect deforming position and gripping point correctly, and could classify socks of same pattern into same group. The study concludes that the method to revise the deformation using multiple sensors is effective.

Keywords: Laundry, House-robot, Socks, Deform.

1 Introduction

Recent years, not only home appliance and consumer electrics, but also house-robots which can carry out the proxy of housework have become popular with the advances in robot technology [1]. According to the definition of house care service in Japan, housework is divided into 5 elements, cleaning, cooking, shopping, washing and tidying up laundry. There are alternative services to each element. For cleaning, room cleaning robot like Roomba has become popular as a housekeeping robot [2]. Window cleaning robot also appeared recently [3]. For cooking, microwave oven is common consumer electronics. For shopping, shopping at online and get delivery by courier have become easier. For washing, washing and drying machine is common too. But there is no alternative for the tidiness of laundry by robots or any consumer electronics (Table 1)[4]. If automation of laundry tidiness is possible, it would save time and reduced burden of the elderly's body.

In this study, we analyzed the functions required to tidy up laundry by robots. As a result, we found that the functions can be divided into four states, the laundry classification, socks pairing and storage process. We thought that the pairing of socks is important, because socks cannot be used if they have not been paired, and pairing

* Corresponding author.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 303–313, 2015.
DOI: 10.1007/978-3-662-46742-8_28

of socks takes a lot of trouble in laundries tidiness. Therefore, we try to develop a system for pairing socks automatically by robots. In this paper, we show the recognition part of this system that includes deforming detection and the similarity determination of socks by the use of image processing and 3D information.

Table 1. Alternative services

Cleaning	Cooking	Shopping	Washing	Tidying
				

2 Laundry State and Action

It is difficult that make robot tidy the laundry. Because various processes are present in the tidiness, and the robot itself is hard to grasp the next required action. Therefore we analyzed the tidiness and defined four states for the sake of simplify the command to the robot. Fig.1 shows the camera image of each state.

-State1: This state is that there is a pile of all kind of laundries (Fig.1a).

-State2: One laundry is separated from the pile of laundry to recognize it (Fig.1b).

-State3: Socks that pairing is completed are arranged (Fig.1c).

-State4: If nothing is in the camera, it is determined on this state (Fig.1d). This represents the completion of all laundry tidiness. This state is the same as the state in which humans tidy the laundry.

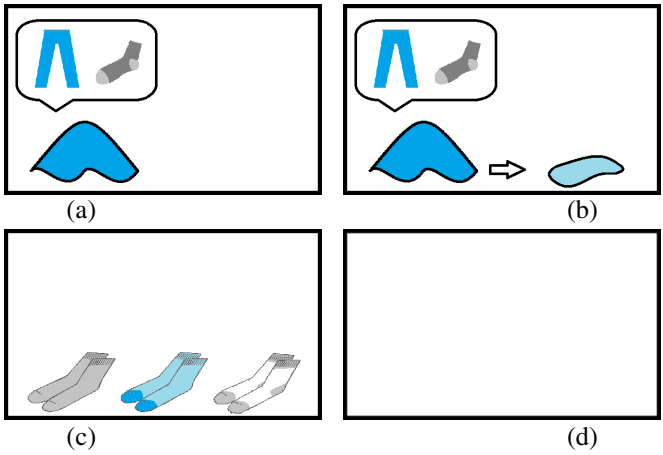


Fig. 1. Image of Laundry State

Table 2 shows the actions to be performed in each state. "Target" represents target that is tracked by the system to take action. "IF" represents the state of the target. "THEN" represents the action of system. "Transition" represents the state after the transition. Fig.2 is a state transition diagram illustrating Table 2.

Table 2. Actions of each state

State	Target	IF	THEN	Transition	
1	pile of laundry	exist	object moving process	state2	1-1
	pile of laundry	none	-	state3	1-2
2	recognized objects	pants	tidying process	state1	2-1
	recognized objects	socks	pairing process	state1	2-2
	recognized objects	none	-	state1	2-3
3	small clothes	exist	tidying process	state3	3-1
	small clothes	none	-	state4	3-2
4	pile of laundries	exist	-	state1	4-1
	recognized objects	exist	-	state2	4-2
	small clothes	exist	-	state3	4-3
	all clothes	none	-	finished	

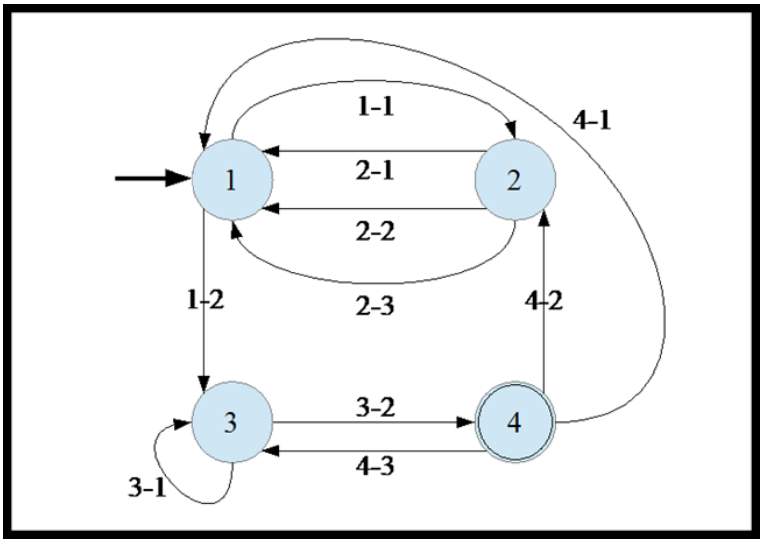


Fig. 2. State transition diagram

3 Proposed Method for Socks Pairing

In case of socks in the second stage, the initial situation is that one sock is separated from the pile of laundry. The final situation is a state that socks are paired by same fabric pattern. In order to achieve the pairing by the system, it is necessary to obtain the fabric pattern and the shape of the object. However, if the deforming and overlapping occurred, it is difficult to obtain accurate feature of socks. Therefore, it is necessary to detect and to revise deforming and overlapping of the socks before pairing. So, we describe methods of the detection and revision of deforming and overlapping, and of the pairing that is performed by determining the color features from obtained image.

We use the Kinect from Microsoft to get RGB image and 3D image, because synchronization of RGB image and 3D image is easy[5]. We have assumed that the system has recognition unit with Kinect and the unit that control small robots with the functions of grip and release of socks and move the ground like Roomba.

3.1 Detection and Revision Function of Socks Deforming

This function detects deforming and overlapping of socks. If deforming and overlapping have occurred in socks, height of the sock becomes more than twice at that portion. Using this, we performed the detection of deforming and overlapping. We show the step of this function.

-Input sock image.
-Get the depth image of sock.
-Get the height difference from depth image (First process).
-Detect straight lines from the height difference, and detect matching portion by comparing lines and outline of sock (Second process).
-Determine the gripping position from detected deformation (Third process).

First process is that detect the height difference by obtaining the depth image of the sock. The sock in the depth image is represented grayscale as shown in Fig.3. Grayscale of the sock vary depending on the distance from the Kinect. We have set the depth recognition range from Kinect and brightness value as shown in Fig.4. Brightness value of the floor is 0. Socks Brightness value of near the floor is 127, and near the Kinect is 0. Because 127 divided by the range of 15 cm, the change of brightness value 1 represents the change of thickness about 1mm of the sock.

Threshold that used for the detection of the difference in height is the average brightness value of the sock. The system detects a part that brightness value is smaller than the threshold as the overlapping part as shown in Fig.5. And determine the height difference from contours of overlapping part (Fig.6).

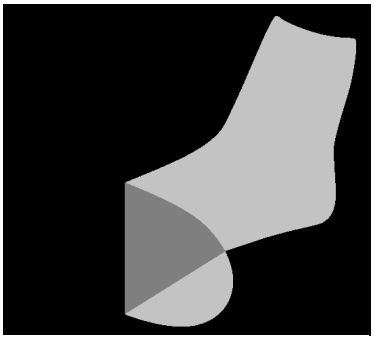


Fig. 3. The sock in the depth image

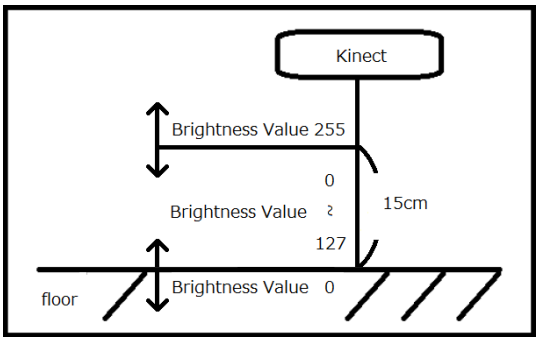


Fig. 4. Recognition range and Brightness value

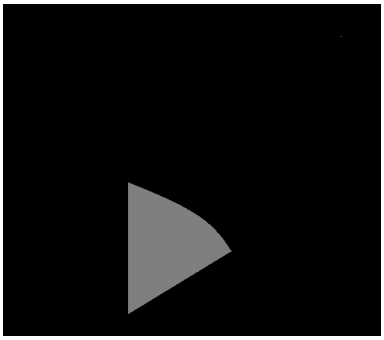


Fig. 5. Overlapping part in sock

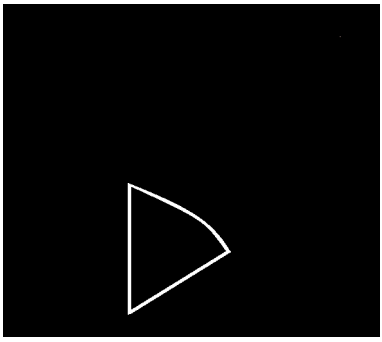


Fig. 6. Height difference image

Second process is to detect the straight line from the previously described image of the height difference. We determine that the detected straight lines are the candidates of overlapping edge (Fig.7). We determine the deforming position by comparing the candidates and outline of socks (Fig.8, 9).

Third process is determination of the gripping position and revision of the deformation. In Fig.10, line segment AB is edge of the deforming position that is decided in the second process. Point C is the midpoint of the line segment AB. We draw a straight line that perpendicular to the line segment AB and passing through the point C. Point D is that this vertical line intersects with the outer peripheral. The system determines the point D as the gripping position (Fig.10). System revises the deformation by moving point D at the position of point symmetry with respect to the point C. If there are multiple deformation, the process is repeated from the left side of the image until deformation disappears.

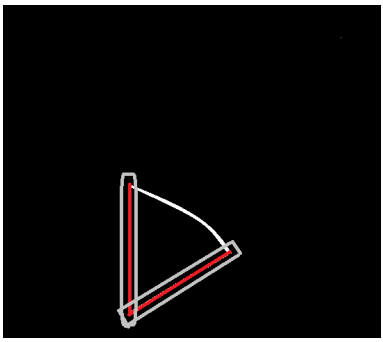


Fig. 7. Candidate of overlapping

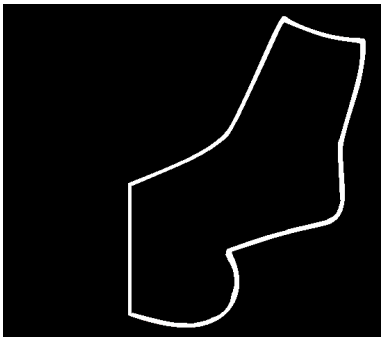


Fig. 8. Outline of socks

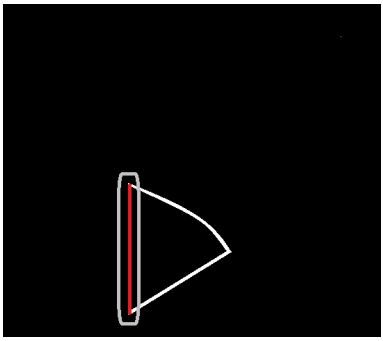


Fig. 9. Deforming point

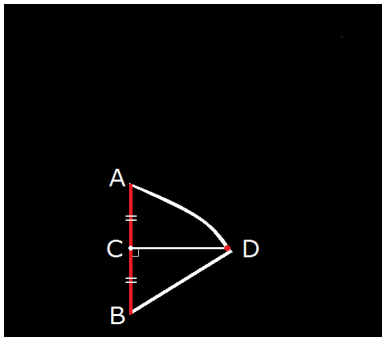


Fig. 10. Third process

3.2 Pairing Function

We create a color histogram by getting RGB image of the sock that is revised the deformation, and we determine the pair of socks by comparing histogram. The system outputs different group number for different socks. If the system determines a sock of same pattern with former one, the system outputs the same group number. The system changes the location of the sock in accordance with the output.

4 Experimental

4.1 Experimental purposes

The purpose of this experiment is to check whether system constructed in this study works as expected. Therefore, we conducted experiments to evaluate judgment of gripping position and socks similarity determination. We evaluated the system by precision, recall and F-measure that are a performance evaluation index in information search.

4.2 Experimental Procedure

We arranged socks and Kinect as shown in Fig.11. We detected deforming of socks and evaluated whether the deformation which was detected agreed with the deformation which has occurred in practice. Next, we evaluated whether can be paired correctly by revision of the deformation and pairing operation. For comparison, we evaluated each pairing of three patterns, only deformed socks, only not deformed socks, and half and half. We performed the experiment 10 times with socks of 10 pairs (20 socks) that have different fabric pattern.

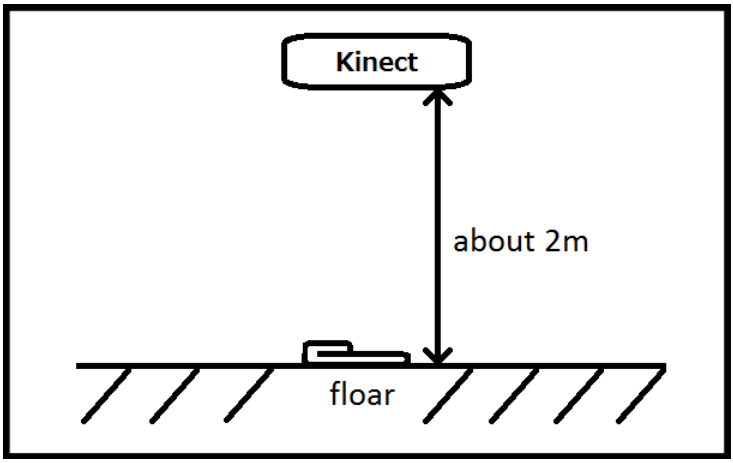


Fig. 11. Experimental configuration

5 Results

Fig.12 shows an example of experimental result of the deformation position and gripping point detection. The average brightness value in Fig.12 example is 98.7, Fig.12b shows the thicker part using this threshold. Fig.12e shows the deforming position. System can detect deforming position correctly. Fig.12f shows the gripping point D. System can detect gripping point correctly.

Table 3 shows the performance evaluation results of deformation detection. "Precision" of table 3 represents the rate of deforming socks per the socks that detected the deforming by the system. "Recall" of table 3 represents the rate of socks that detected the deforming by the system per the all deforming socks. “F-measure” of table 3 shows the performance of the deformation detection determined from the Precision and Recall.

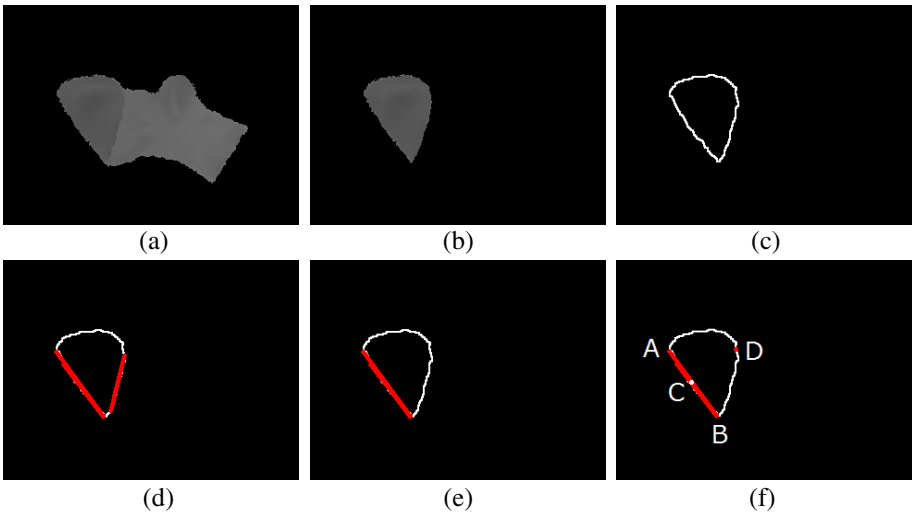


Fig. 12. Example of the result of the deformation position detection experiment

Table 3. Performance evaluation of deformation detection

Precision	Recall	F-measure
0.85	0.74	0.79

Table 4 - table 8 shows an example of the result of pairing experiment. The sock images of table 4 are entered into the system. Graphs of table 5 – table 8 represent the rate(%) of each color in the socks. The vertical axis of the graph represents the rate of color value, and the horizontal axis represents color value 0-255. “Output” shows the determination result of the system. System can classify socks of same pattern into same group.

Table 4. Image of input sock

Sock1	Sock2	Sock3	Sock4
			

Table 5. Graph and output of socks1

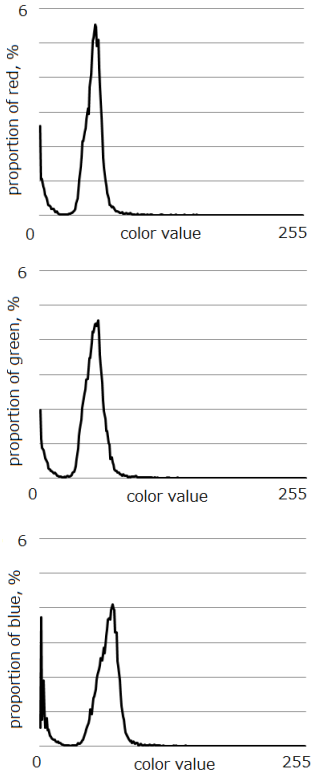
Each color graphs	Out-put
	1

Table 6. Graph and output of socks2

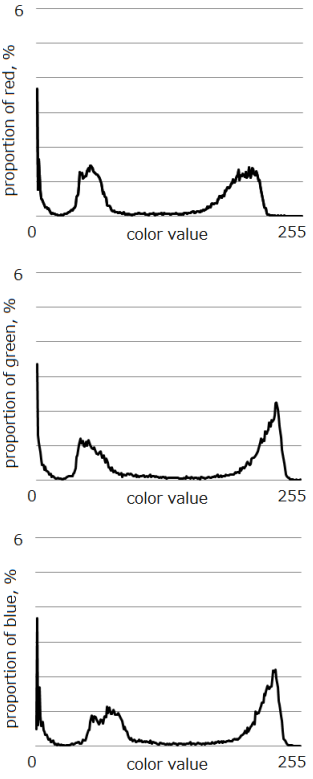
Each color graphs	Out-put
	2

Table 7. Graph and output of socks3

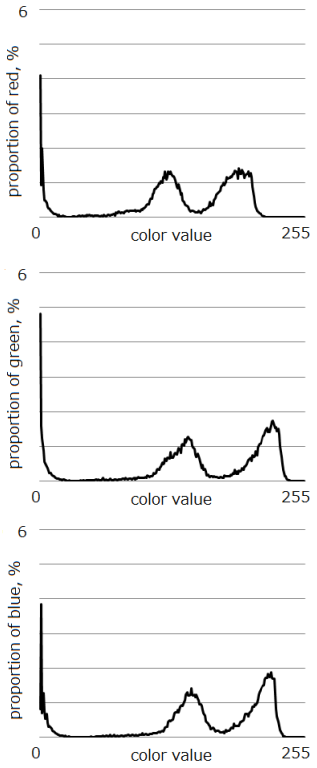
Each color graphs	Output
	3

Table 8. Graph and output of socks4

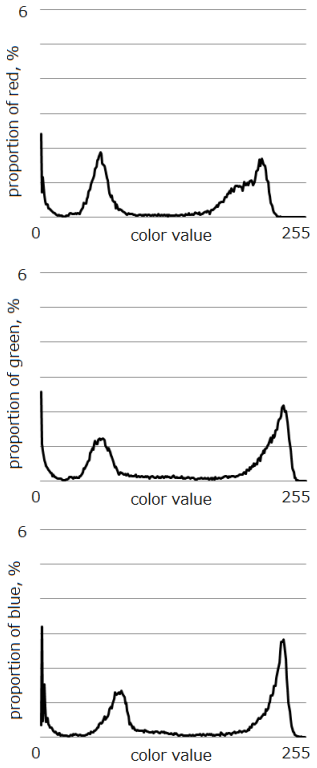
Each color graphs	Output
	2

Table 9 shows the performance evaluation results of pairing. "Precision" of table 9 represents the rate of socks paired correctly per the socks paired by the system. "Recall" of table 9 represents the rate of socks paired by the system per the all pair socks. "F-measure" of table 9 shows the performance of the pairing function determined from the Precision and Recall.

Table 9. Performance evaluation results of pairing

Sock state	Precision	Recall	F-measure
deformed	0.24	0.26	0.25
half and half	0.37	0.42	0.39
not deformed	0.69	0.73	0.71

6 Conclusions

In this paper, we have developed a system for deformation detection and the socks pairing by using image processing and 3D information. By deformation detection process, recognition rate of deformation detection was 74% and precision rate of the recognized deformation was 85%. From this result, deformation detection by this system is effective. In addition, from the results of pairing experiment, the accuracy of the whole system will be around 70%. Socks that deformation was not accurately detected were something thin like stockings in deformation position detection experiment. This is conceivable that was judged as "no deformation" because height difference was not able to detect.

From pairing result, F-measure of the socks pairing that are not deforming was about three times the socks that are deforming, because it is possible to detect such patterns that are hidden when it is deforming.

As challenges for the future, we try to improve the matching method in socks pairing by recognition of design pattern, socks size, etc.

References

1. Okada, K., Ogura, T., Haneda, A., Fujimoto, J., Gravot, F., Inaba, M.: Humanoid motion generation system on HRP2-JSK for daily life environment. In: Proc. ICMA 2005, vol. 4, pp. 1772–1777. IEEE (2005)
2. iRobot Corporation, Roomba, http://www.irobot.com/
3. ECOVACS, WINBOT, http://www.ecovacs-japan.com/
4. Osawa, F., Seki, H., Kamiya, Y.: Laundering Tasks Performed by a Robot: Spreading Out and Sorting of the Clothes, vol. 70(10) (2004)
5. Torgerson, E., Paul, F.W.: Vision-guided robotic fabric manipulation for apparel manufacturing. In: Proceedings of the 1987 IEEE International Conference on Robotics and Automation, vol. 2, p. 1196 (1987)
6. Sakamoto, D., Honda, K., Inami, M., Igarashi, T.: Sketch and Run: A Stroke-based Interface for Home Robots. In: Proceedings of the 27th International Conference on Human Factors in Conputing Systems, CHI 2009, Boston, USA, April 4-9, pp. 197–200 (2009)
7. Microsoft, Kinect, http://www.xbox.com/en-US/xbox-360/
8. Liu, F., Tang, J., Zhao, R., Tang, Z.: Abnormal behavior recognition system for ATM monitoring by RGB-D camera. In: Proceedings of the 20th ACM International Conference on Multimedia, p. 1295 (2012)
9. Sun, C., Bao, B.-K., Xu, C., Mei, T.: Kinect-based visual communication system. In: Proceedings of the 4th International Conference on Internet Multimedia Computing and Service, p. 55 (2012)
10. Panger, G.: Kinect in the kitchen testing depth camera interactions in practical home environments. In: Proceedings of the CHI 2012 Extended Abstracts on Human Factors in Computing Systems, p. 1985 (2012)

Design and Implementation of Skeletonization

Kartika Gunadi[1], Liliana[1], and Gideon Simon[2]

[1] Informatics Department of Petra Christian University
Siwalankerto 121-131, Surabaya
{kgunadi,lilian}@petra.ac.id
[2] Informatics Department of Petra Christian University
Siwalankerto 121-131, Surabaya
poinkpoink91@hotmail.com

Abstract. The development of image processing technology is well developed today, such as object recognition in an images. Algorithm to obtain accurate results in object recognition continues to be developed. One example is the handwriting recognition application. This application is usually used to archive records or documents from physical form such as a notebook or a letter in the form of files digital. One of the initial process in image processing is image segmentation, and one of the methods used is skeletonizaton.

This Skeletonization uses Discrete Local Symmetry. The process begins by setting the active contour of the image. From the active contour triangulation process is done. And from the process the symmetry points are defined. Skeletonization process is then performed using point symmetry obtained.

The results show that the size of an image greatly affect the outcome of the process of skeletonization using Discrete Local Symmetry. Discrete Local Symmetry methods suitable for ribbon-like shaped objects.

Keywords: Image Processing, Skeletonization, Discrete Local Symmetry.

1 Introduction

The development of image processing technology is developing at this time, including the introduction of the object in the image. Algorithm to obtain accurate results in object recognition continue to be developed. One example is a handwriting recognition application. These applications are typically used to archive records or documents from a physical form such as a notebook or a letter in the form of digital files.

Object recognition in images requires several stages in the process of image processing one of which is the process of image segmentation. Many methods can be used in image segmentation process, one of which is Skeletonization. Skeletonization process is used because the results are given in the form of the image frame that still has the original characteristics of the topology and shape of the image. Where the results of the image frame can be used further in image processing applications such as the application of pattern recognition, handwriting recognition and fingerprint recognition [2].

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 314–322, 2015.
DOI: 10.1007/978-3-662-46742-8_29

Many existing skeletonization methods being developed. One example is using the method of thinning skeletonization. But the method that is mentioned to have weaknesses where the result of that process is affected by noise. So the result order obtained do not comply with the characteristics and the original form of the image. In addition, the size of the image also affect the length of time the process because the number of pixels the skeletonization will in the process more and more. To overcome these problems, a skeletonization method developed robust against noise using the method of Discrete Local Symmetry.[4]

2 Skeletonization

Skeletonization is one of image processing which has enough global functions. Skeletonization is widely developed because the framework has a good structure and still have the characteristic shape of the image making it suitable for use in image processing applications, one of which is pattern recognition.

Many methods can be used in the process of using operators such as skeletonization morphologi e.g. thinning and (called pruning). And some other methods like using level sets, curve evolution, constrained delaunay triangulation [3] and distance transform.

In general the method in the skeletonization can be divided into two pixel-based and non-pixel-based. In pixel-based method used is usually thinning or distance transforms. While in a non-pixel-based, only the contours of the pixels of an object that is used to process the skeletonization. [4]

Method of non-pixel-based usually has a faster processing time than the pixel-based because fewer processed data. To get an order of an object using contours, discrete local symmetry of the object should be obtained accurately. Discrete local symmetry can be calculated from the pixel of the objects contour. Techniques for identifying discrete local symmetry matching used in low resolution image or medium, this is because in the high resolution image, all the pixels of the image contours used in the calculations. [4]

2.1 Discrete Local Symmetry

Discrete local symmetry (DSL) describes a symmetry between the contour of the pixels of an image with contour segments between two adjacent contour pixels. The concept was developed to obtain a the symmetry between the contour pixel and contour segment that showed a series of contour pixel [5]. A contour is divided into a series of segments with the linearization process where two ends of each contour segments connected by straight line segments.

It is assumed that no straight line segments intersect each other at interior points. Thus, the contours of the shape of the picture can be indicated with a planar straight line graph G (V, E) where ends of E and vertices V is a straight line segment and the end of the contour. A vertex v ∈ V and point e ∈ E form a generalized Discrete Local Symmetry (gDLS) if loop in the triangle formed by a T v and e there was no tip from

the vertices of G which is visible from all other vertices of t. and T is contained in the object in the original image. Two dots are visible from every segment of the line that connects them not intersect with point e ' ∈ e. A gDLS will be DSL if in the main segment of the contour, there are only two and only 2 contours pixels.

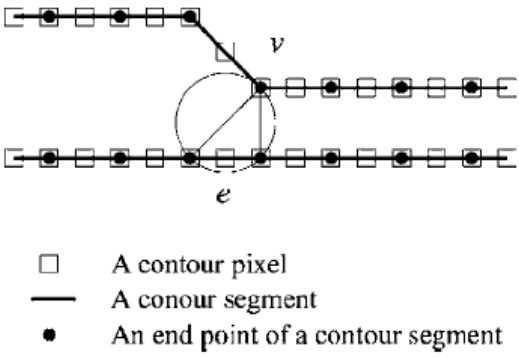


Fig. 1. Metode Discrete Local Symmetry

2.2 Contour Pixel

Contour Pixel is the set of pixels that form the boundary edges of the Contour can be open or closed [1]. Regional boundaries are useful to describe the shape of the object in the image analysis stage. Contours can be divided into two, namely, closed contour and open contour. Close contour is contour with the boundaries that surround an area, and open contour can be line or part of the boundaries of the area.

3 System Design

The software system developed for the process of Skeletonization method using Discrete Local Symmetry is composed of five main processes. The first is the process of processed imagery grayscaling, where from the image will be transformed into the image of colored grayish. The second is the process of thresholding the image results from the images obtained so grayscaling consisting only of black and white. This is done to help the next process to get accurate results. Then proceed with the process of with Edge Detection. In this process the pixels of the image of the contour is obtained.

The fourth process is the process of Triangulation, where contour pixel Edge Detection result of the process divided into triangles. The last process is the search axis of symmetry derived from the process of Triangulation process results and Skeletonization are displayed.

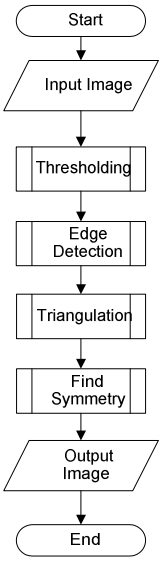


Fig. 2. The Main Process Flowchart

3.1 Triangulation

The process of Triangulation is the main process in the method of Discrete Local Symmetry. On the process of triangulation is the result of the process of Edge Detection will be divided into active contour which would have made the point to draw a triangle. Description of the process in general are shown in Figure 3.

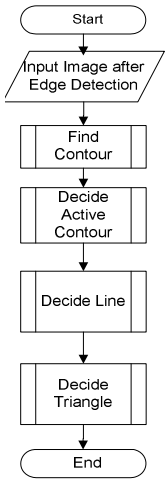


Fig. 3. Flowchart Triangulation

4 Implementation and Testing

In this section it will be shown the test results using the method of Discrete Skeletonization application Local Symmetry. The test will be divided into four parts, namely the validation test, test on paper, test on the image and test the thickness.

4.1 Test Validation

In this section the results of the use of the application in the validation test. To test the validity of application use image by comparing image desired results with images thickened. From the results of the validation test the original image and the result still has the same basic shape.

Source Image	Bolded Image	Result
		
		
		

Fig. 4. Validation Test Results

4.2 Test on Writing

In this chapter will compared the results with the other 2 methods of application which is thinning and Distance Transform of writing.

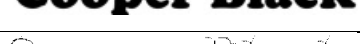
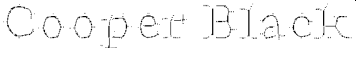
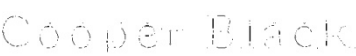
Source Image	
Thinning Methods	
Distance Transform Methods	
Discrete Local Symmetry Methods	

Fig. 5. Results of test on Writing the alphabet

Source Image	
Thinning Methods	
Distance Transform Methods	
Discrete Local Symmetry Methods	

Fig. 6. Test results on Chinese Writing

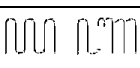
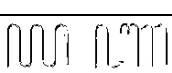
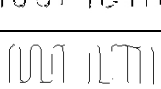
Source Image	
Thinning Methods	
Distance Transform Methods	
Discrete Local Symmetry Methods	

Fig. 7. Results of test on Writing Java

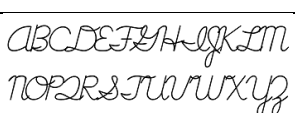
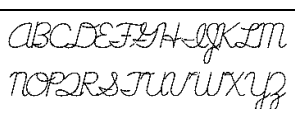
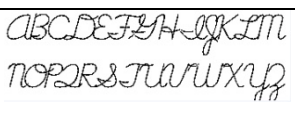
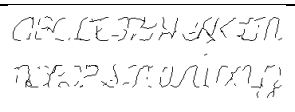
Source Image	
Thinning Methods	
Distance Transform Methods	
Discrete Local Symmetry Methods	

Fig. 8. Test result on Handwriting

From the test results in Figure 5 and Figure 6 shows that the basic form of writing is still legible, but drop out in some parts of the letter. In Figure 7 a thin piece of writing to be lost so that the basic form of writing changed. Whereas in Figure 8 results out of shape. Loss of yield due skeletonization posts are too thin.

4.3 Test in Figure

In this section the results will be compared with two other methods of application are thinning and Distance Transform on the image.

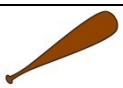
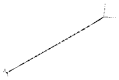
Source Image	
Thinning Methods	
Distance Transform Methods	
Discrete Local Symmetry Methods	

Fig. 9. Test Results in Figure Sticks

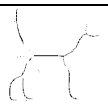
Source Image	
Thinning Methods	
Distance Transform Methods	
Discrete Local Symmetry Methods	

Fig. 10. Test results on the image of the dog

From testing on a stick and the silhouette of a dog, the skeleton of the method of Discrete Local Symmetry still has the same basic form with the original image.

4.4 Thickness Test

On the sub chapter applications will be tested against the thickness of various sizes in units of pixels.

Size	24 *px*	36 *px*	48 *px*
Image Source			
Result			
Size	72 *px*	96 *px*	120 *px*
Image Source			
Result			

Fig. 11. Test Results Thickness in letter

Size	3 *pt*	5 *pt*	7 *pt*
Image Source			
Result			
Size	9 *pt*	11 *pt*	13 *pt*
Image Source			
Result			

Fig. 12. Test Results The thickness of the image circle

From the test against the thickness of this proves that the method is weak against the thin image. In Figure 11 the new form of writing seen in size 36 pixel 12 Pictures while on the basic forms of the new circle is visible on the thickness of the 5 points, although the results are not perfect.

5 Conclusion and Advise

5.1 Conclusion

Based on the test results can be concluded as follows:

- Through testing we can conclude that different active contour placement in an image affect the outcome of the Local Discrete Symmetry method. This is because the placement of which is based on active contour pixels to make the results of different triangulation process. This can be seen in which the thickness of the test results of the skeleton of the letter G in Figure 11 should be the same.
- Based on the test results thickness Local Discrete Symmetry method is not suitable for thin images because it can eliminate the basic form or interruption of the skeleton of the image results.
- From the test results Local Discrete Symmetry method is more suitable for ribbon-shaped objects like such in writing and written Mandarin alphabet.

5.2 Advise

The few things that can serve as suggestions in the process of further development, among others:

- Development of applications in order to process the skeleton for the posts with a small size.
- Development of algorithms in order to find the midpoint of an image accurate.
- Development of algorithms in order to avoid breaking the result of skeletonization process.

References

1. Munir, R.: Kontur dan Representasinya, p. 141 (2005), http://informatika.stei.itb.ac.id/~rinaldi.munir/Buku/Pengolahan%20Citra%20Digital/Bab-9_Kontur%20dan%20Representasinya.pdf (retrieved April 26, 2013)
2. Zhao, F., Tang, X.: Preprocessing and postprocessing for skeleton-based fingerprint. Pattern Recognition 40, 1270–1281 (2007)
3. Zou, J.J., Yan, H.: Skeletonization of Ribbon-Like Shapes Based on Regularity and Singularity Analyses. School of Engineering and Industrial Design University of Western Sydney, Australia (2001)
4. Zou, J.J.: A Fast Skeletonization Method. School of Engineering and Industrial Design University of Western Sydney, Australia (2003)
5. Zou, J.J.: Efficient Skeletonisation Based on Generalised Discrete Local Symmetries. Optical Engineering 45(7), article number 077205, pp. 077205-1–077205-7 (2006)

A Computer-Aided Diagnosis System for Vitiligo Assessment: A Segmentation Algorithm

Arfika Nurhudatiana

School of Information Systems, Bina Nusantara University,
Jalan Hang Lekir 1 No. 6, Jakarta Indonesia 10270
anurhudatiana@binus.edu

Abstract. Vitiligo is a condition where depigmentation occurs in parts of the skin. Although vitiligo affects only 0.5% to 1% of the world's population, vitiligo may impact patients' quality of life greatly. The success of vitiligo treatment is influenced by the accuracy of vitiligo assessment method used for monitoring the treatment progress. A popular method for assessing vitiligo is the Vitiligo Area Scoring Index (VASI), where VASI score is calculated based on the visually observed degree of depigmentation in the skin and the area of depigmented skin. This approach has two limitations. First, it has inter-examiner variability, and second, its accuracy can be very low. While the first limitation can be addressed through training, the second limitation is more difficult to address. With visual observation, positive but small progress may be unnoticed since it usually takes some time to notice skin color changes manually. To overcome this limitation, we propose in this paper a vitiligo lesion segmentation algorithm as a part of a computer-aided diagnosis system for vitiligo assessment. The algorithm reads color skin image as input and makes use of the Fuzzy C-Means clustering algorithm and YCbCr and RGB color spaces to separate pigmented and depigmented skin regions in the image. The corresponding degree of depigmentation is then estimated. The algorithm was evaluated on low resolution Internet images of skin with vitiligo. The results are encouraging, indicating that the proposed algorithm has a potential for use in clinical and teledermatology applications.

Keywords: Segmentation, clustering, computer-aided diagnosis, medical image analysis, pigmentation.

1 Introduction

Vitiligo is a skin disorder where pale or white patches develop due to lack or absence of melanocytes. In vitiligo patients, the melanocytes are either damaged or destroyed, causing the absence of melanin (pigments) in the affected areas of the skin. Vitiligo affects around 0.5% to 1% of the world's population and it may have profound psychological impact on patients' quality of life. Although vitiligo patients do not usually feel ill, the appearance of the vitiligo can be very distressing, especially for the darker-skinned subjects where the white patches are more noticeable. In addition, vitiligo

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 323–331, 2015.
DOI: 10.1007/978-3-662-46742-8_30

treatments may take a long time. This may cause vitiligo patients to suffer low morale. Some patients may even quit treatments after going through some disappointments from not seeing any progress.

The success of vitiligo treatment is influenced by the accuracy of vitiligo assessment method used for monitoring the treatment progress. Alghamdi et al. [1] reviewed and classified various vitiligo assessment techniques based on three criteria: 1) as subjective, semi-objective or objective, 2) as microscopic or macroscopic, and 3) as based on morphometry or colorimetry. Subjective methods include clinical evaluation by a dermatologist and the vitiligo disease activity score (VIDA), which is a six-point scale scoring method ranging from +4 to -1 based on dermatologist's own opinion of the present disease activity over time [2]. Wood's lamp, which shines ultraviolet (UV) light on the skin, is commonly used in the visual examination of vitiligo. Semi-objective methods include Vitiligo Area Scoring Index (VASI) [3] and point-counting methods. VASI score is calculated based on two parameters, which are the degree of depigmentation and the surface area affected by the disease. Objective methods include software-based image analysis and other assessments using special imaging devices such as tristimulus colorimetry, spectrophotometry and confocal laser microscopy (CLM). It was reported in the study that VASI and Wood's lamp are some of the best techniques available for assessing the degree of pigmentary lesions and measuring the extent and progression of vitiligo [1].

Although VASI method is considered to be one of the best vitiligo assessment methods, calculating VASI score based on visual observation has limitations. The first limitation is the inter-examiner variability among different dermatologists/clinicians, and the second limitation is the low precision of manual observation. Since VASI score estimates degree of depigmentation into small categories of percentage values (i.e., 25%, 50%, and 75%) and a relatively longer period of time is typically required to notice changes in the skin color, minor pigmentary changes may remain unnoticed for a period of time until a larger area of the skin changes. While the first limitation can be addressed through training, the second limitation is more difficult to address. We present in this paper a preliminary work on a computer-aided diagnosis system to assess vitiligo lesions based on VASI. The system consists of an algorithm which reads RGB image as input, segments skin from the background, separates depigmented skin and normal skin into two clusters, and then calculates the degree of depigmentation in the image.

The rest of this paper is organized as follows. Section 2 presents related work on vitiligo lesion segmentation. Section 3 presents the Vitiligo Area Scoring Index (VASI). Section 4 presents the proposed algorithm. Section 5 gives the experimental results and discussion. Finally, Section 6 provides conclusion and future work.

2 Existing Work

Several computational methods have been proposed for vitiligo lesions assessment. Nugroho et al. [4-6] proposed to use principal component analysis (PCA) and independent component analysis (ICA) for vitiligo lesion segmentation, where images of

skin were separated into melanin and hemoglobin spaces. The algorithm was evaluated on small vitiligo regions of several patients.

In addition to the work in [4-6], Aydin et al. [7] compared point counting and digital planimetry techniques for estimating vitiligo surface area. Their method was semi-automatic since observers were involved in their techniques.

Our algorithm is different from the techniques in [4-7] since our algorithm is fully automatic and is targeted to specific body sites instead of small vitiligo regions. In addition, our algorithm is targeted on low to medium resolution images which can be produced by using general smartphone cameras.

3 Vitiligo Area Scoring Index (VASI)

Vitiligo Area Scoring Index (VASI) is a quantitative parametric scoring method introduced by Hamzavi et al. [3]. The concept of VASI scoring method is derived from the Psoriasis Area and Severity Index (PASI) scoring method used in psoriasis assessment [8]. VASI score is calculated as follows:

$$\text{VASI}= \sum_{\text{AllBodySites}} [\text{HandUnits}] \text{ x} [\text{Depigmentation}], \tag{1}$$

where [HandUnits] refers to the body surface area affected by vitiligo in hand units and [Depigmentation] refers to the degree of depigmentation within each hand unit measured patch.

One hand unit covers the area of palm plus the volar surface of all the digits (i.e., fingers) and is approximately 1% of the total body surface area [9]. The degree of depigmentation within each hand unit measured patch is estimated to the nearest of the following percentages: 100%, 90%, 75%, 50%, 25%, 10%, and 0%.

The body is divided into five body sites (i.e., hands, feet, trunk, upper extremities excluding hands, and lower extremities excluding feet) and the product calculation is applied to each body site. For example, given a body site, if there are 2 hand units of surface area with 50% depigmentation, the VASI score for that body site is 2 hand units x 0.5 depigmentation = 1%. The final VASI score is then calculated from all body sites (Eq. 1).

4 The Proposed Algorithm

The YCbCr color space and the Fuzzy C-Means (FCM) algorithm, which is a clustering algorithm, are utilized to perform vitiligo lesion segmentation. The YCbCr color space is used because the high Cr coordinates of the CbCr two-dimensional plane in the YCbCr color space localizes skin color very well. The FCM algorithm is used since pixels of skin tend to form homogeneous clusters in image and thus the skin segmentation can be approached as a clustering problem.

The proposed algorithm is illustrated in Fig. 1. The algorithm is composed of two steps. The first step is the skin segmentation step, which serves as preprocessing to

localize the skin regions in the image, and the second step is the vitiligo lesion segmentation step. The skin segmentation step works as follows. Given an input RGB image, the image is first transformed into the YCbCr color space for skin segmentation. The Cr channel is then extracted and smoothed using a mean filter. The smoothed channel is then segmented into two different clusters using the Fuzzy C-Means (FCM) clustering algorithm [10]. Since skin tends to have high intensity values in the Cr channel, the cluster centers are initialized with 10 and 90 percentiles of the intensity values in the channel, where the first cluster represents the background and the second cluster represents the skin. This process produces skin mask M (see Fig. 2).

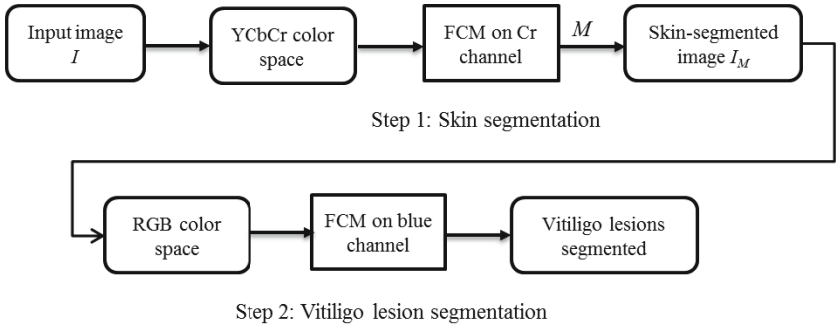


Fig. 1. Illustration of the proposed algorithm

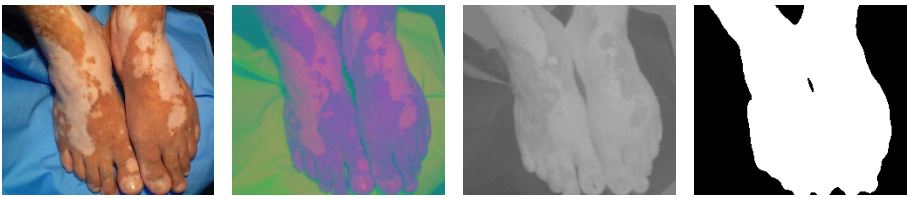


Fig. 2. Illustration of the skin segmentation process; (a) input RGB image, (b) the input image (a) in YCbCr color space, (c) Cr channel extracted from (b), and (d) the skin mask *M*

The skin mask M is then smoothed using a morphological operator which removes all connected components with sizes smaller than 10% of the image pixel resolution. The skin mask M is applied to the input image to obtain skin-segmented image I_M, where the values of the pixels belonging to the first cluster (background) are set to zero and the values of the pixels belonging to the second cluster (skin) are retained.

After skin segmentation, the skin-segmented image goes through vitiligo lesion segmentation process. In this step, a different color channel, which is the blue channel of the RGB color space, is used. Blue channel wavelengths are close to the ultraviolet

wavelengths, while the wavelengths of the red and green channels penetrate into the deeper layers of the skin. Therefore, skin is more sensitive to the blue channel wavelengths compared to the red channel and green channel wavelengths. Fig. 3 compares the appearance of vitiligo skin in red, green, and blue channels. It can be observed that the contrast difference between the normal skin and the depigmented skin is highest in the blue channel (Fig. 3(c)) compared to the rest of the channels (Fig. 3(a)-(b)).

(a) (b) (c)

Fig. 3. Comparison of vitiligo skin appearance in (a) red channel, (b) green channel, and (c) blue channel

The same FCM clustering operation with three clusters is then applied to the skin-segmented image. The resulting clusters are then further categorized based on the centroid values. The cluster with centroid almost equal to zero represents the black background and thus is labeled with black color in the skin map, while the remaining two clusters with non-zero centroids represent the skin. The cluster with lower centroid value represents pixels of pigmented skin and thus is labeled with gray color in the skin map, while the cluster with higher centroid value represents pixels of depigmented skin and thus is labeled with white color in the skin map.

After the vitiligo lesion segmentation is finished, the degree of depigmentation in that body site is estimated based on the two generated clusters. In this preliminary study, the vitiligo images were downloaded from the Internet and the dpi (dot-per-inch) resolutions of the images are unknown. Therefore, the number of pixels representing one hand unit cannot be estimated and thus the VASI score cannot be properly calculated. To overcome this, the degree of depigmentation in this study is calculated as the number of pixels in the depigmented skin cluster divided by the total number of skin pixels, which is the total number of pixels in both the depigmented skin cluster and the normal skin cluster.

$$\text{Depigmentation} = \frac{\#\text{ pixels depigmented}}{(\#\text{ pixels depigmented} + \#\text{ pixels normal skin})}. \quad (2)$$

5 Experimental Results and Discussion

The algorithm was evaluated using publicly available vitiligo images downloaded from the Internet [11]-[13]. The pixel resolutions of the images ranged from 240x180 to 640x480 pixels. These pixel resolutions are considered as low and medium resolu-

tion. The dpi resolutions were unknown since no actual scale information of the body sites (in inch or centimeter unit) was provided. The other vitiligo lesion segmentation algorithms were evaluated on images with higher resolutions [4-6]. These images showed smaller areas of the skin (not the full body site) and were captured using DSLR camera. Since the existing algorithms were not designed for low resolution images, they are not applicable for comparison.

Fig. 4 shows the results of the proposed algorithm on vitiligo skin images from feet, legs, hands, and arms body sites. It can be observed that the pigmented and the depigmented regions are well separated in all the images. The segmentation result is slightly rough on the sides of the arms in Fig. 4(d) where hair is present. However, this is not a problem since it does not affect (or at least only slightly affects) the degree of depigmentation calculation. It can also be observed that fingernails and toenails are still included in the cluster of depigmented skin. For these images, the actual degree of depigmentation needs to be adjusted from the original calculation given by Eq. 2. However, the adjustment value depends on the scale of the image, which actually varies for the dataset used in this study. This variation can be minimized if clinical images are used. Clinical images are usually captured in standardized image acquisition condition where background, pose, and scale can be completely controlled. Pose and background can be controlled by asking the users to stand against the wall in a standard pose then place their hands and feet on the examination bed in certain pose (see Fig. 4). Scale can be controlled by setting a fixed distance between the camera and the object (e.g., between the clinician's desk and the wall) and using the same zoom factor for the camera. A ruler can be placed near to the object (e.g., hand) so that precise dpi information can be calculated from each image.

In terms of speed, the algorithm provides real-time response if it is to be used during face-to-face consultation in clinic or self-assessment by patients. The algorithm was implemented using Matlab on a computer laptop with 2.40 GHz Intel processor. It required 7-10 seconds to process each input image. Since Matlab uses interpreted language, the Matlab code is executed without first compiling it into machine-language instructions. The speed would be faster if the algorithm was written in low level languages (compiled code) such as C++.

Since different subjects have different skin tones, the performance of the algorithm may vary in different subjects. Some subjects have fair and light skin tone (e.g., Chinese subjects) while some other subjects have dark skin tone (e.g., Negro subjects). Vitiligo lesion segmentation is more challenging in fair-skinned subjects due to the lower contrast between pigmented and depigmented skin. An additional contrast enhancement method may be required for preprocessing the image before vitiligo lesion segmentation can be performed.

Although with this algorithm vitiligo lesion assessment can be performed automatically, it should be noted that using a computer-aided system for vitiligo assessment does not necessarily mean that the assessment is solely performed by the software. The proposed system serves as a tool to provide more detailed information on the current condition of the patient. It is important that the final judgment is still made by the attending dermatologist/clinician.

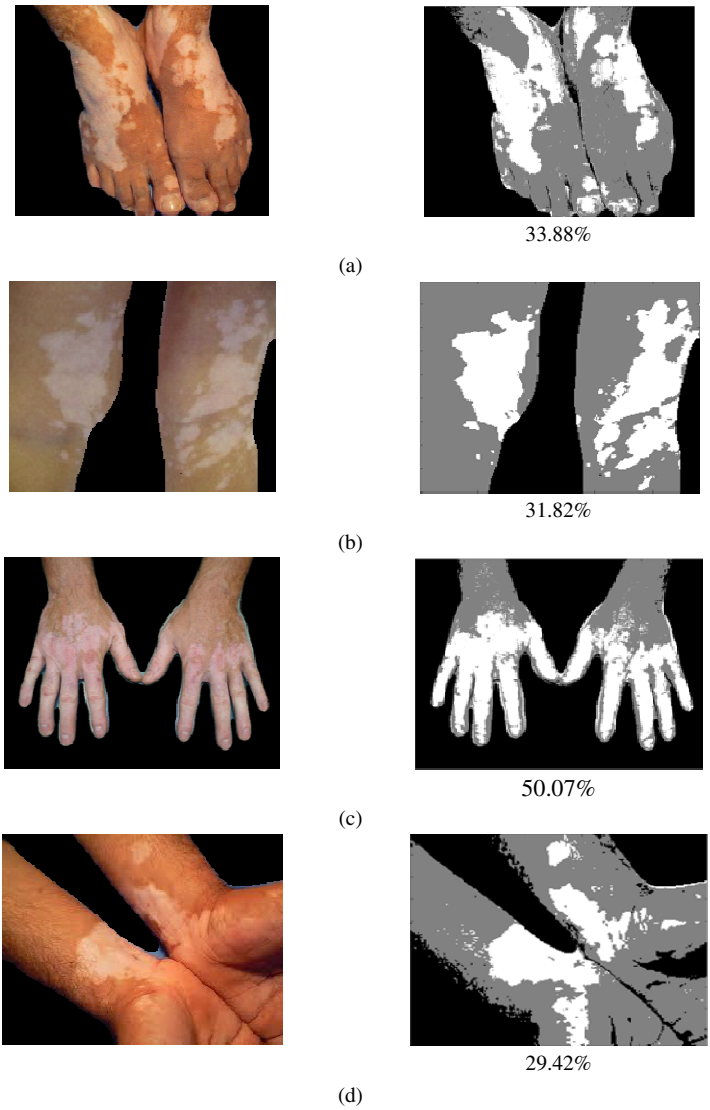


Fig. 4. Vitiligo lesion segmentation results; (a) feet, (b) legs, (c) hands, (d) arms; the black, grey, and white colors in the skin map indicate background, normal skin, and depigmented skin respectively. The percentage indicates the estimated degree of depigmentation in the skin image.

6 Conclusion and Future Work

We present in this paper a preliminary work on a computer-aided diagnosis system for the assessment of vitiligo. The proposed vitiligo lesion segmentation algorithm was evaluated on Internet images of vitiligo skin in various body parts. The results show that the proposed algorithm works well for the images. The whole processing is

considerably fast, indicating that the algorithm has a potential to be implemented in clinical application. It is also applicable for teledermatology application since it takes low resolution image as input and thus image acquisition can easily be performed using smartphone cameras.

In this preliminary study, the hand unit for calculating VASI score could not be properly estimated since there was no information on the dpi resolution of the hand and other body sites. We plan to collect images from various body sites with a standard scale, pose, and background setting. A ruler will be positioned near to the skin to be able to estimate the dpi resolution of each body site. This way, the number of pixels equivalent to one hand unit can be properly estimated and the whole process of VASI score calculation can be fully automated. The adjustment values for removing toenails and fingernails from the degree of depigmentation calculation can also be accurately estimated. The algorithm will also be improved to make it more robust to varying skin conditions including lighter skin tone and skin with dense hairs.

Skin covers the human body in three-dimension. A possible extension of the current study is to utilize the three-dimensional imaging system for faster and more accurate skin image acquisition. This future work will also include the extension of the algorithm on three-dimensional images.

Acknowledgments. The author would like to thank the medical doctors and researchers who have provided their insights and contribution to this work.

References

1. Alghamdi, K.M., Kumar, A., Taïeb, A., Ezzedine, K.: Assessment Methods for the Evaluation of Vitiligo. J. Eur. Acad. Dermatol. Venereol 26(12), 1436–1471 (2012)
2. Njoo, M.D., Das, P.K., Bos, J.D., Westerhof, W.: Association of the Koebner Phenomenon with Disease Activity and Therapeutic Responsiveness in Vitiligo Vulgaris. Arch. Dermatol. 135, 407–413 (1999)
3. Hamzavi, I., Jain, H., McLean, D., Shapiro, J., Zeng, H., Lui, H.: Parametric Modeling of Narrowband UV-B Phototherapy for Vitiligo, Using a Novel Quantitative Tool: The Vitiligo Area Scoring Index. Archives of Dermatology 140(6), 677–683 (2004)
4. Nugroho, H., Fadzil, M.H.A., Yap, V.V., Norashikin, S., Suraiya, H.H.: Determination of Skin Repigmentation Progression. In: Proc. IEEE Int. Conf. Engineering in Medicine and Biology Society, pp. 3442–3445 (2007)
5. Ahmad Fadzil, M.H., Norashikin, S., Suraiya, H.H., Nugroho, H.: Independent Component Analysis for Assessing Therapeutic Response in Vitiligo Skin Disorder. Journal of Medical Engineering & Technology 33(2), 101–109 (2009)
6. Nugroho, H., Ahmad Fadzil, M.H., Shamsudin, N., Hussein, S.H.: Computerised Image Analysis of Vitiligo Lesion: Evaluation Using Manually Defined Lesion Areas. Skin Research and Technology 19, 72–77 (2013)
7. Aydin, F., Senturk, N., Sahin, B., Bek, Y., Yuksel, E.P., Turanli, A.Y.: A Practical Method for the Estimation of Vitiligo Surface Area: A Comparison Between the Point Counting and Digital Planimetry Techniques. Eur. J.Dermatol. 17(1), 30–32 (2007)
8. Fredriksson, T., Pettersson, U.: Severe Psoriasis-Oral Therapy with a New Retinoid. Dermatologica 157(4), 238–244 (1978)

9. Rossiter, N.D., Chapman, P., Haywood, I.A.: How Big is a Hand? Burns 22(3), 230–231 (1996)
10. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
11. Patterns of Depigmentation in Vitiligo, Vitiligo Zone, http://www.vitiligozone.com/research-studies/patterns-depigmentation-vitiligo
12. Vitiligo, American Osteopathic College of Dermatology, http://www.aocd.org/?page=Vitiligo
13. Vitiligo Support International, http://www.vitiligosupport.org/faq.cfm

Face Recognition for Additional Security at Parking Place

Semuil Tjiharjadi and William Setiadarma

Computer Engineering Department, Maranatha Christian University
Jl. Suria Sumantri 65 Bandung - 40164
semuiltj@gmail.com

Abstract. As a country that has a lot of car users, the parking management and its security are quite complicated issues. This paper is described the use of facial recognition system using Eigenface method, to identify the drivers of four-wheeled vehicles. the face of the driver of the car was taken while taking a parking ticket and will be recorded as comparative data source that will be used to identify the driver when paying the parking fee. The results of this research tests performed are able to recognize when the driver out, but the results will be better if the driver's face image capture using multiple cameras so Eigenface face recognition system has enough resources to recognize the driver's face. This system is expected to reduce the number of car thefts committed in parking places.

Keywords: face recognition, eigenface, car park.

1 Introduction

Data processing has been performed by humans for a long time. Humans also find mechanic and electronic equipments to help people in the calculation and data processing, in order to obtain faster results. Computers is a long evolution of human inventions since ancient times, in the form of mechanical or electronic appliances.

Nowadays computers and supporting tools have been included in every aspect of life and work. Computers are possess more than just the usual mathematical calculations, including the computer system in a supermarket cashier who is able to read the code of the goods shopping, telephone exchange that handles millions of calls and communications, computer networks and the Internet that connects various places in the world. Many things are easier to use now.

Computer application have shifted from regular computing applications to intelligence system. One of the concept of intelligence is how to program a computer to be able to recognize a person's face by using a camera. This paper describe the concept to recognize persons's face and improve the security in the parking area.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 332–347, 2015.
DOI: 10.1007/978-3-662-46742-8_31

2 Concepts

2.1 Hardware Design

The driver will press the entry switch when the car get into the parking. It will provide the input signal to the microcontroller. The microcontroller will forward the input signal to computer device. Digital cameras take drivers pictures. Another signal was given to the printer to print the serial number of parking ticket as the identity of the car while in the parking area. When the car leave the parking area and stop at the exit gate then operator will scan the ticket number and once again the camera will take the driver picture and compare it with the database. When the system can recognize the driver then the system will tell the operator. On the other hand, if the system can not recognize the driver then computer will give warning to operator.

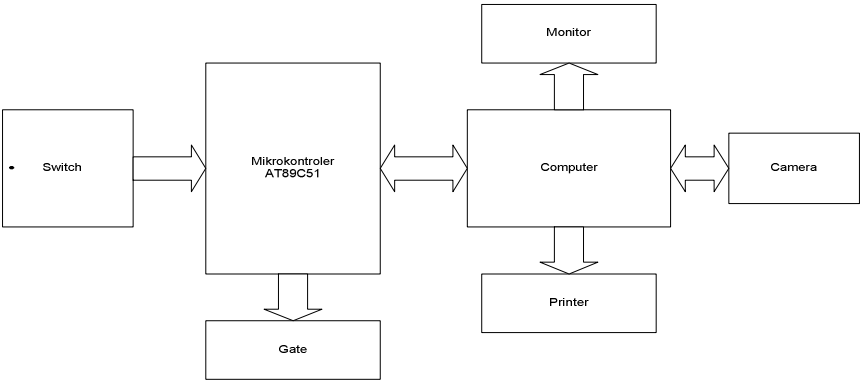


Fig. 1. Face Recognition System For More Secure at Parking area

2.2 Digital Image

Digital image is an image that can be processed by a computer. For example, in Picture 2, a size of 80x80 pixel grayscale image was taken in part (small box) size 6x6 pixels. Thus, the monitor will display a small box. However, that is stored in the computer's memory is just pictures that showed great intensity in each pixel.

2.3 Image Processing

Digital image processing is a form of processing information with an image as input, such as drawing a picture or frame of the video. The output is an image that can be used as a function of the image itself.

A digital image A (m, n) has been described in a two-dimensional plane analogue that was obtained from an analog image A (x, y) on a continuous two-dimensional

plane of the sampling process each period that has been digitized. A continuous two-dimensional image (x, y) is divided into N rows and M columns, the point of intersection are both referred to as pixels.

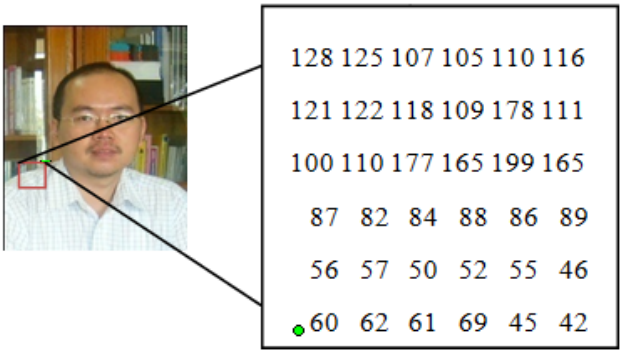


Fig. 2. 80x80 pixel grayscale digital image

Digital images are numerical data that can be processed by a computer to obtain information, since the digital image is represented in a matrix. Operations on the digital image are basically manipulating matrix elements. Matrix elements can be manipulated single element (pixel), a set of adjacent elements, or all the elements of the matrix. An assortment of image processing operations such as:

- Geometric transformation such as zoom in, zoom out, and rotate images.
- Improving colours like dark, light and contrast adjustment, or convert into a form different colors.
- Merging two or more images.
- Segmentation image
- Editing and improve digital images and others

2.4 Face Recognition

In general, the image of a face recognition system is divided into two types, namely systems feature-based and image-based system. In the first system to use the features extracted from the image components of the face (eyes, nose, mouth, etc.) And then the relationships between these features are modelled geometrically. While both systems use the raw information from the pixel image which is then represented in a specific method, such as Principal Component Analysis (PCA) which is then used for classification of image identity.

2.5 Principal Component Analysis (PCA)

PCA procedure is basically aimed at simplifying the observed variables by means of shrinking (reducing) dimensions. This is done by removing the correlation between

independent variables through the transformation of the independent variable to the origin of a new variable that is not correlated at all or commonly referred to as principal component

Fig. 3. Differences in color and grayscale images

2.6 Eigenface Algorithm

Eigenface is one of face recognition algorithm based on Principal Component Analysis (PCA). To produce Eigenface, large collection of digital images of human faces taken in the same lighting conditions and then normalized and then processed at the same resolution (eg m x n), and then treated as mn -dimensional vector whose components are taken from the value of its pixels.

Eigenface face recognition algorithm steps are : [5]

1. Flat Vector Preparation

The first step is to develop a training image into one single matrix. For example, the stored image size H x W pixels and the number N of fruit, it will have flat vector with dimensions N x (H x W).

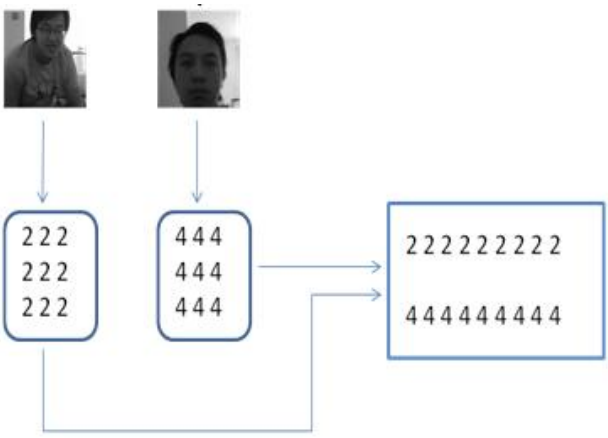


Fig. 4. Sample Preparation Flat Vector [2]

2. Flat Vector average
When flat vector obtained, add the whole row in order to obtain a single matrix measuring 1 x (H x W). After that the matrix N was the number of images to obtain Mean flat Vector

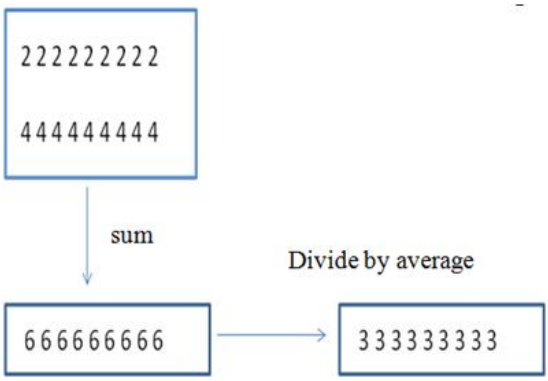


Fig. 5. Example average count of Flat Vector [2]

3. Determine the value of eigenface
By using the average will be calculated eigenface flat vector for flat vector matrix that has been compiled. How to reduce the rows of the matrix with the average flat vector. If the score is below zero, change the value to zero.

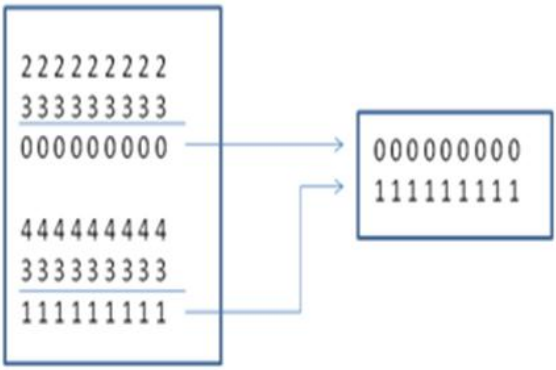


Fig. 6. Eigenface value [4]

4. Identification process
If the given image to be identified (test face), then the identification step is as follows: Eigenface value calculation for test face matrix, in the same way with the determination eigenface to flat vector. [6]

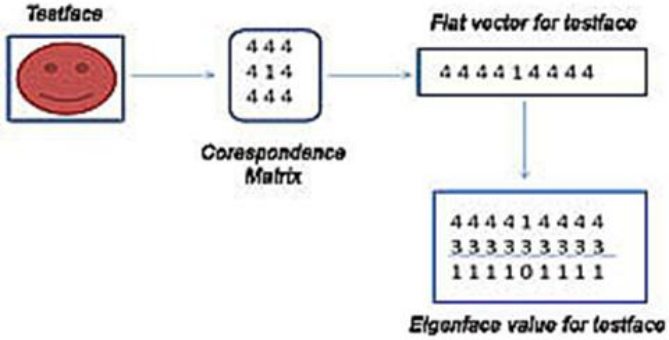


Fig. 7. Example calculation Eigenface value for face recognition

After Eigenface value for test face is obtained, it can be identified by determining the distance (distance) with the shortest eigenvector of Eigenface from training images. Firstly, specify the absolute value of the use of row i of the matrix Eigenface training image with Eigenface of test face, then add the constituent elements of the vector resulting from the reduction of the distance d earlier and found the index i. It will do for all the lines and find the smallest value d. [7]

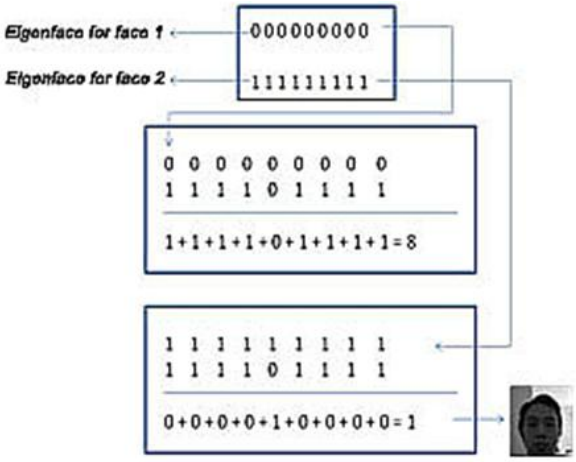


Fig. 8. Process of Identification

3 Design

3.1 Block System Diagram

Note block diagram:

- The face image is captured using a webcam. Results of capturing this is the type .bmp image files.

- The face image is then normalized with several stages. First, the image quality derived color to grayscale. The size of the face image is also uniform, to a size of 80 x 80 pixels.
- Having obtained normalized facial image, calculate the eigenvalues of the face image, such that the value of x.
- In the face image database contains a collection of face images. From this collection of images each of its eigenvalues calculated and collected in a vector which we call eigenvector. Suppose obtained values (x1, x2, x3, ... xn).
- Matching process is done by matching the value of x with the values of the Eigenvector and find the closest value.
- If the value that is closest to it is found, looking face data corresponding to the last value and show the image of the face and facial image data. [8]

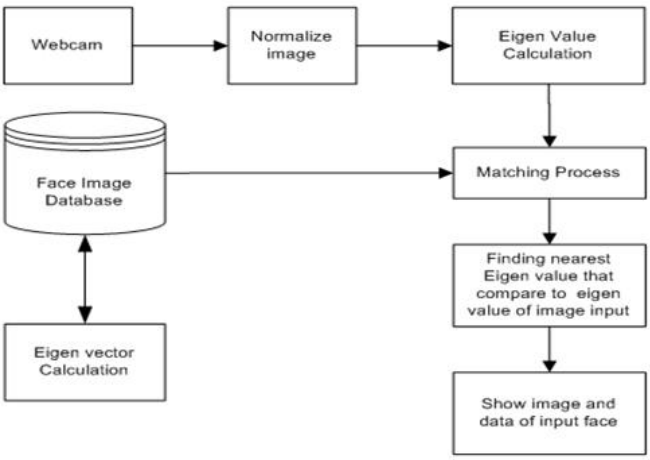


Fig. 9. Block Diagram

3.2 Database

The database consists of table's id to store the data from a face image and setting the table for storing the value of the nearest percentage when recognizing faces. In the table, the primary key is the id with the AutoNumber data type as a place to store the index that is used to index the face image storage. ID number and name of the text data type is used to store the data from a face image. [3]

Table 1. Database ID

Field Name	Data Type	Field Size
Id*	AutoNumber	-
Nrp	Text	10
Nama	Text	50

In the table setting, there is no primary key because it will contain the data one alone. The data will be filled by MinEigen and the value will contain the values to percentages in recognizing faces. The value can be filled through the program.

Table 2. Database Setting

Field Name	Data Type	Field Size
Nama	Text	50
Nilai	Number	LongInteger

Data access between Microsoft Access and Visual Basic 6, uses ActiveX Data Object (ADO) as the data consumers who use data and OLE DB as the data providers that accommodate and unload the data. The relationship between the application and database components can be seen in Figure 10 and 11.

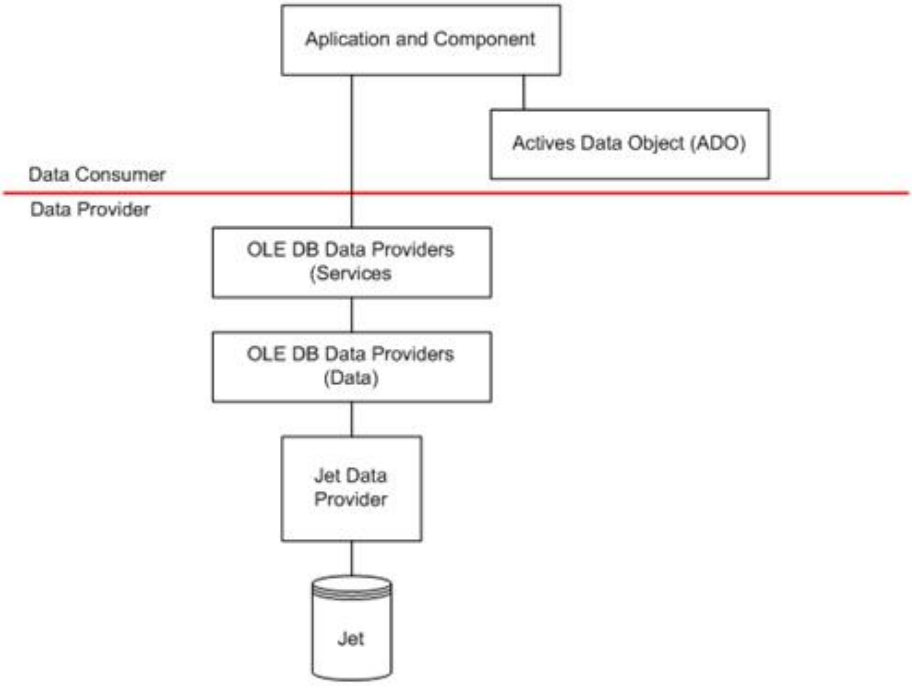


Fig. 10. Consumer data and Provider

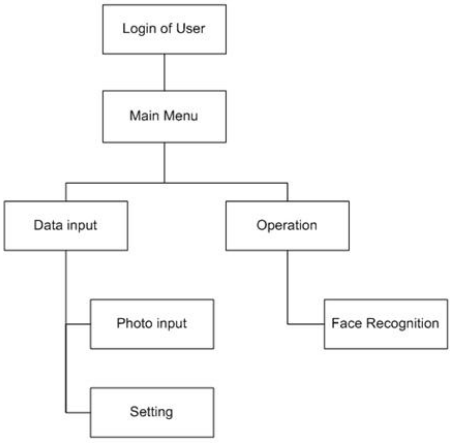


Fig. 11. Program Structure

3.3 Program Structure

Menu structure used on facial recognition program can be seen in the following chart:

User Login : first face recognition.
Main menu : Source and data to leave.
Input Data : use to input data.

There are 2 options submenu, namely Input Photos, used to take photos for the facial image and data entry. The second is setting, that is used to change the face recognition similarity percentage. [9]

3.4 Input Design

Design input is designed as a user interface where the user can enter either the data input via the keyboard and via webcam. The design input for face recognition program consists of the parts:

1. Form Login User Design
   In this form the user is quite simply pressing the button Login to access a menu of facial recognition program.
2. Form Input Photo Design
   In this form the user can take a picture and fill the face image data of the face image. The component used to capture the image is ezVidcap and to display the database is DataGrid.
3. Form Setting Design
   In this form may be a change in the degree of similarity is desired for face recognition.

Fig. 12. Form Login Design

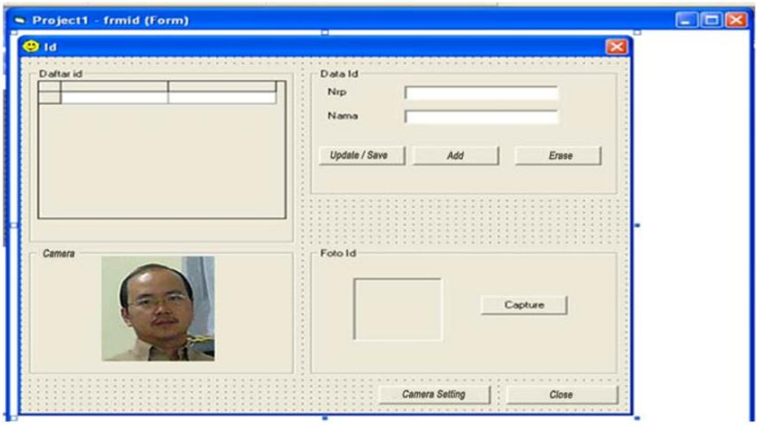


Fig. 13. Form Photo Input Design

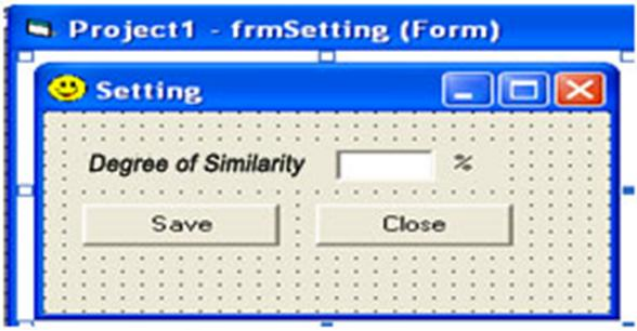


Fig. 14. Form Setting Design

4. Form Face Recognition Design

In this form the face image captured from a webcam to a computerized then matched with a face image in a database and data. Face recognition process carried out in this form

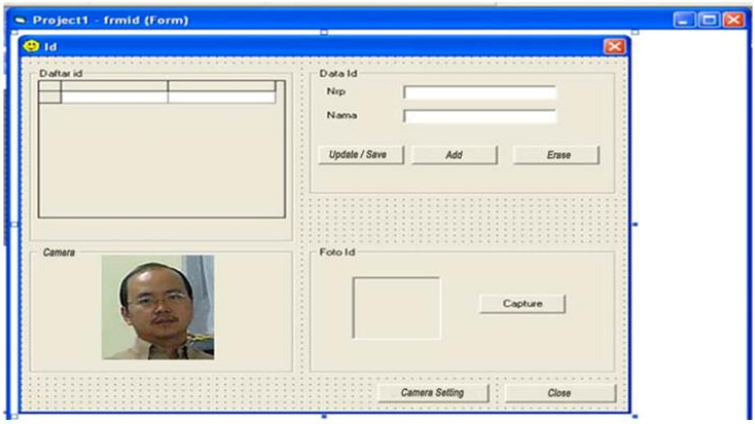


Fig. 15. Face Recognition Form Design

The data Flowchart of face recognition program is divided into three parts.

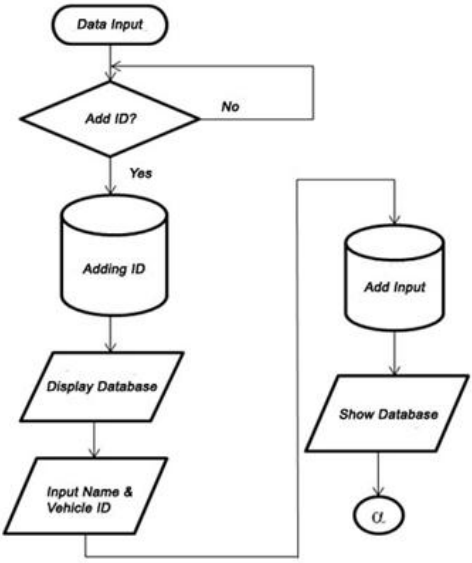


Fig. 16. Flowchart of Data input

In the process of data input, must first be adding the id first. This is done so that the id of the database by type AutoNumber increases. After being shown on the database and the user can fill Name input and vehicle ID will be stored in the database then the data will be displayed on the form interface.

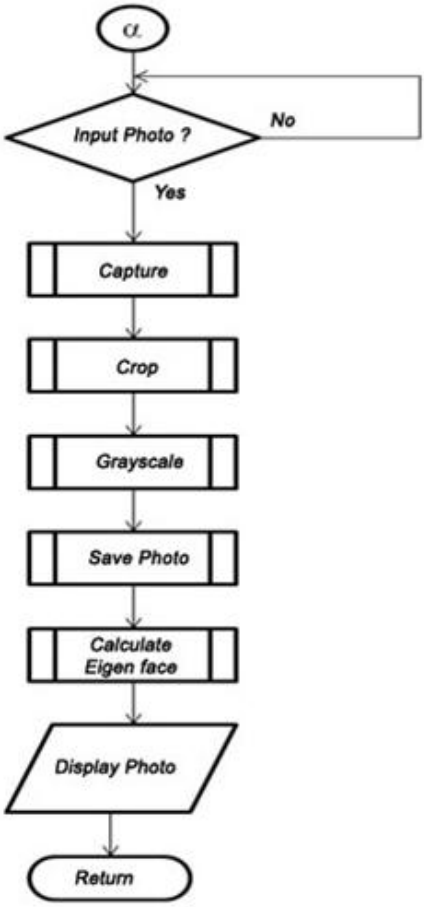


Fig. 17. Flowchart of Photo Input

In the photo input process, the user will be prompted to capture the image of the user's face. Once captured, the face image obtained will be processed in size to 80 x 80 pixels, lowered into a grayscale image quality, is stored in the database, and will be calculated value Eigenface her. The photo is displayed on the form interface.

In the process of face recognition, face's image will be captured then the obtained image will be processed in size to 80 x 80 pixels, lowered quality of the image to grayscale, and then the image will be calculated matching its eigenface value and will look for the most value eigenface approached with a face image which is stored in the database. [10]

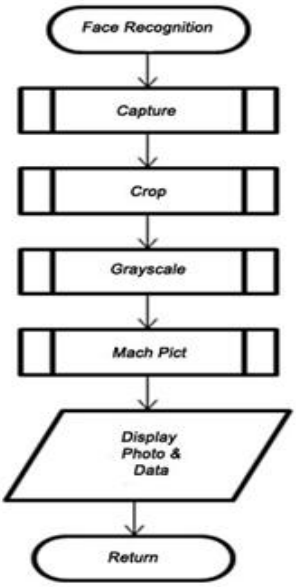


Fig. 18. Face Recognition

4 Testing

In the process of face recognition, face to be recognized will be captured then the image obtained will be processed in size of 80 x 80 pixels, lowered quality of the image to grayscale, and then the image will be calculated matching its Eigenface value and will look for the most Eigenface value approached with a face image which is stored in the database.

The first observation of data, based on the results that have been obtained of experiments with the same background, but there are other objects in the background, bright lighting, and captured face without expression or formal. Experiments were carried out on 5 people as the drivers, each of them were performed the experiment 5 times.

Table 3. Observation 1

	1	2	3	4	5	6	7	8	9	10
Martha	√	√	√	√	√	√	√	√	√	√
Rina	√	√	√	X	√	√	√	√	√	√
Oscar	√	√	√	√	X	√	X	√	√	√
Insan	√	√	√	√	√	√	√	√	√	√
Willy	√	√	√	√	√	√	√	√	√	√

Note :
√ = indentified face X = not recogniced face

From the first experiment, evidence of failure as much as 8 times. Face is not identified in accordance with the face you want to identify. The success of face recognition gained only 68%.

2nd observation data obtained based on the results of experiments with the same background, but there are other objects in the background, bright lighting, and captured face can not have expression. Experiments were carried out on 5 people; each person only has 5 facial databases and performed the experiment as much as 10 times.

Table 4. Observation 2

	1	2	3	4	5	6	7	8	9	10
Martha	√	√	√	√	√	√	√	√	√	√
Rina	√	√	√	X	√	√	√	√	√	√
Oscar	√	√	√	√	X	√	X	√	√	√
Insan	√	√	√	√	√	√	√	√	√	√
Willy	√	√	√	√	√	√	√	√	√	√

Of the 2nd experiment, evidence of failure as much as 3 times. Face is not identified in accordance with the face you want to identify. The success of face recognition gained through 94%.

3rd observation data obtained based on the results of experiments with the same background is blue, bright lighting, and facial expression captured. Experiments were carried out on 5 people; each person has a database of face as much as 1 face and conducted the experiment as much as 5 times per person. The results are as follows:

Table 5. Observation 3

	1	2	3	4	5
Willy	√	√	√	√	√
Oscar	√	X	√	X	√
Rina	√	√	X	X	√
Martha	√	√	√	X	√
Insan	√	√	X	√	√

Table 6. Observation 4

	1	2	3	4	5	6	7	8	9	10
Martha	√	√	√	√	√	√	√	√	√	√
Rina	√	√	√	√	√	√	√	√	√	√
Oscar	√	√	√	√	X	√	X	√	√	√
Insan	√	√	√	√	√	√	√	√	√	√
Willy	√	√	√	√	√	√	√	√	√	√

From experiment 3, evidence of failure as much as 6 times. Face is not identified in accordance with the face you want to identify. The success of face recognition obtained up to 76%.

4th observation data obtained based on the results of experiments with the same background is blue, bright lighting, and facial expression capture. Experiments were carried out on 5 people, each person has a face database and do as much as 5 faces trial as much as 10 times per person. The results are as follows:

Of the 2nd experiment, evidence of failure as much as 2 times. Face is not identified in accordance with the face you want to identify. Accuracy in face recognition algorithm using Eigenface on the 2nd experiment obtained up to 96%.

5 Conclusion

From the results of experiments that have been conducted, it can be concluded as follows:

1. Face recognition applications can be created using eigenface algorithm for face recognition, software Visual Basic, and Microsoft Access as a database of images image.
2. Eigenface algorithm works by calculating the average pixel of the images that have been stored in a database. The average pixel value will be obtained eigenface of each image and then will look for the nearest eigenface value of drawing a face image you want to identify.
3. The face of someone who can be identified is stored in the database face images that have been processed by the algorithm eigenface, otherwise it would not be recognizable face or faces are recognizable not appropriate.
4. The success of face recognition algorithm using eigenface will be higher if the face image database stored more and more. This conclusion is derived from the results of the comparison between the first observational data that were attempted in 5 people with background who have other objects, but each person has only one face image database. The 2nd data observations attempted to 5 people with a background that has other objects but each there are 5 people face image database. 2nd observational data with a face image databases more, have a higher success rate than the first observational data that there is only one database only. As well as a comparison between the data observations 3rd attempted in 5 people with background, there are no other objects but each person has only one face image database. The 4th data observation was attempted in 5 people with the background there are other things, but every man has 5 face image database. 4th observational data with a face image databases more have a higher success rate compared with observational data 3rd there is only one database only.

References

1. Fatta, H.A.: Rekayasa Sistem Pengenalan Wajah. C.V Andi Offset, Yogyakarta (2009)
2. Madcoms.: Seri Panduan Pemrograman Database Visual Basic 6.0 Dengan Crystal Report. C.V Andi Offset, Yogyakarta (2002)
3. Mesran.: Visual Basic. Mitra Wacana Media, Jakarta (2009)
4. Sutoyo, T., Mulyanto, E.: Teori Pengolahan Citra Digital. C.V Andi Offset, Yogyakarta (2009)
5. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience (1991)
6. Davies, Ellis, Shepherd: Perceiving and Remembering Faces. Academic Press, London (1981)
7. Kirby, M., Sirovich, L.: Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Transactions on Pattern analysis and Machine Intelligence (1990)
8. Heseltine, T., Pears, N., Austin, J., Chen, Z.: Face Recognition: A Comparison of Appearance-Based Approaches. In: Proc. VIIth Digital Image Computing: Techniques and Applications, vol. 1 (2003)
9. Moghaddam, B., Wahid, W., Pentland, A.: Beyond eigenfaces: probabilistic matching for face recognition. In: Third {IEEE} International Conference on Automatic Face and Gesture Recognition (1998)
10. Belhumeur, P.N., Hespanha, Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection (1997)

Optic Disc Segmentation Based on Red Channel Retinal Fundus Images

K.Z. Widhia Oktoeberza, Hanung Adi Nugroho, and Teguh Bharata Adji

Department of Electrical Engineering and Information Technology, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia
widhia.oktoeberzakz.mti13@mail.ugm.ac.id,
adinugroho@ugm.ac.id, adji@ugm.ac.id

Abstract. Glaucoma is a one of the serious diseases that occurs in retina. Early detection of glaucoma can prevent patients from blindness. One of the techniques to support the diagnosis of glaucoma is developed through the detection and segmentation of optic disc area. Optic disc area is also useful in assisting automated detection of abnormalities in the case of diabetic retinopathy. In this work, extracted red channel of colour retinal fundus images is used. Median filter is used to reduce noises in the red channel image. Segmentation of optic disc is conducted based on morphological operation. DRISHTI-GS dataset is used in this research works. Results indicate that the proposed method can achieve an accuracy of 94.546% in segmenting the optic disc.

Keywords: Optic disc segmentation, Red channel, Median filter, Thresholding, Morphological operation, Retinal fundus image.

1 Introduction

Glaucoma and diabetic retinopathy are common eye disease that causes of blindness. Glaucoma is the second causes of blindness in the world. Early detection of glaucoma is important to prevent of blindness [1], [2]. Diabetic retinopathy is a chronic retinal related disease as along-term effect of diabetes. Diabetic retinopathy occurs when retinal vessels are damaged as a result of increasing of glucose circulation in the blood vessels [3]. Patients of glaucoma and diabetic retinopathy increase every year [4], [5] and more than half of them experience blindness. It is a big issue that is experienced by ophthalmologists. Ophthalmologists use retinal fundus images that are obtained from fundus camera to diagnose some of retinal related diseases like diabetic retinopathy and glaucoma with appeared main signs indicated from optic disc area and the presence of pathologies such as exudates and cotton wool spots [6].

Optic disc is a brightness area in retinal as entrance and exit spot of blood vessel and retinal nerve fibers. Centre of optic disc is called cup region [7]. Optic disc plays an important rule in development of automated method for diagnosis of diabetic retinopathy. Other ophthalmic pathologies is called glaucoma also can diagnosis by optic disc segmenting [8], [9]. On the other hand, detection of optic disc region can

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 348–359, 2015.
DOI: 10.1007/978-3-662-46742-8_32

reduce false positive when conducted detection other bright abnormalities like exudates caused by characteristics similarity of them [7]. Figure 1 shows abnormalities in the optic disc area indicated by its abnormal size.

Fig. 1. A digital fundus image showing optic disc area [10]

Many researchers have conducted research on the optic disc detection for diagnosis of glaucoma [1], [4] and diabetic retinopathy [5], [6]. Pachiyappan *et al*. [1] segmented optic disc area by using low pas filtering and thresholding. It is followed by exudates detection, blood vessels extraction and microaneurysms segmentation to classify severity of glaucoma and diabetic retinopathy. Grading severity of diabetic retinopathy can also be done by analyzing FAZ area [11].

Kavitha and Malathi [4] used combination of K-Means clustering and gabor filtering to segment optic disc area for diagnosis of glaucoma. Segmented optic disc is processed by using morphological operation to obtain better result.

Ponnaiah *et al*. [7] applied genetic algorithm (GA) in blue channel image to increase accuracy of detection and removal optic disc area. This proposed method can locate optic disc region very fast and achieved good accuracy. Weighted error rate (WER) was used to evaluate the result.

In [12] used combination of morphology and Kirsch's template to remove optic disk and retinal blood vessel. It was useful for easier exudates detection. Aquino *et al*. [8] located optic disc area by using maximum variance method based on morphology. It was performed in green channel to obtain candidate optic disc pixel (ODP). Circular Hough Transformation was used to segment circular optic disc boundary. It was conducted in red and green channel, the best result of them is selected. Blood vessel was eliminated by rotating linear structuring element to improve efficiency of

optic disc segmentation. Experiment is performed by using MESSIDOR dataset. OD area can localize with achieved percentage 99% and 86% for segmenting.

Godse *et al.* [9] focused on localisation of optic disc. Threshold estimation was applied based on green channel histogram. This step was conducted to find all of bright region that called cluster. Then two criteria, namely area criterion and density criterion are applied on that cluster. Area criterion useful to select candidate area, whereas density criterion selected satisfies area. Threshold was main determinant selected area.

Eadgahi and Pourreza [13] considered maximum entropy for candidate optic disc area. Sobel operator was used to edge detection. Proposed method can perform in lower quality image and tested by using DIARETDB1 dataset. Abbadi and Al-Saadi [14] determined center of optic disc by computing ratio of two channel, namely red and green channel. Optic disc area was located by determining radius from center in four directions.

Reza *et al.* [6] proposed preprocessing in original fundus image by applying average filter to handle variation intensity. Contrast stretching transformation was used to separate bright region from background. Processed image is transformed by using negative transform to obtain fixed threshold as a parameter. Finally, watershed transformation was conducting to distinguish optic disc, exudates and cotton wool spots from background image.

Ranamuka and Meegama [15] eliminated optic disc based on thresholding technique and applied Nilblack's method to obtain binary image. Fang *et al.* [16] first conducted vessel segmentation then applied matched filtering to get center of optic disc. Center and size of optic disc was adjusted by using double ring filter.

The focus of this research work is to identify optic disc area using red channel of retinal fundus images. The contrast of blue channel is low contrast and does not contain much useful information [17], [14], while green channel is good to detect blood vessels [14]. Filtering is applied in preprocessing to make optic disc area clearer. The rest of this paper is organised as follows. Section 2 describes the approach as the proposed method, which is used in this research. Section 3 shows the results and discussion. Conclusion is presented in section 4.

2 Approach

The approach in this research is described as follows. Firstly, original colour fundus image containing three different channels, i.e. red, green and blue channel is extracted. The red channel is used as the input image. Median filtering is applied in preprocessing step to remove noises and to obtain optic disc area be clearly. Otsu thresholding is used to segment optic disc area. For better segmented area, opening operation based on mathematical morphology is applied. Ground truth image is used to validate accuracy of the result. Feature extraction is conducted to know significant feature that can representative optic disc area. Flowchart of the approach is shown in Figure 2.

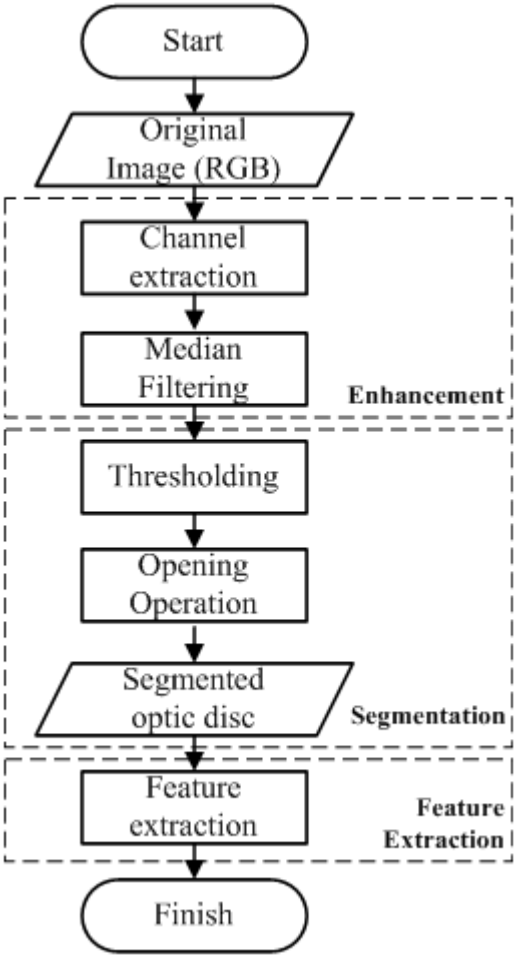


Fig. 2. Flowchart of the approach

2.1 Preprocessing

Channel extracted is conducted to obtain red channel as input image. Noises that appeared in input image are removed by using median filtering. Median filter has a good performance to reduce noise.

Median filter worked by replacing pixel value with median level intensity of adjacent pixels. Pixel value in point (x, y) is entered to median computation [18]. Median filtering is mathematical formulated in following equation.

$$f(x,y) = \underset{(s,t)\in S_{xy}}{median}\{g(s,t)\} \tag{1}$$

2.2 Segmentation

Centre point of image segmentation is thresholding. Pixel values that smaller than threshold value are assumed as background, while others are assumed as object point [18], [19]. Threshold process resulting binary images that consist of two gray level namely black and white. In mathematical formulated is explained in following equation.

$$b(x,y) = \begin{cases} 1, if\ f(x,y) \geq T \\ 0, if\ f(x,y) < T \end{cases} \quad (2)$$

T is threshold intensity. Quality of binary image is very determined by thresholding [20].

Otsu calculated automatically threshold value T based on input images. This method determined a variable that can separate between two even more objects. It is called discriminant analysis. Discriminant analysis maximized that variable to separate object and background images.

A threshold value automatically selected by Otsu [17]. Firstly, Otsu calculated intensity value i of histogram with mathematical formulation in Equation (3). N as number of pixel images while n_i as number of pixels with i intensity. Weighted of object and background is declared in Equation (4) and (5).

$$p(i) = \frac{n_i}{N}, p(i) \geq 0, \sum_{1}^{256} p(i) = 1 \quad (3)$$

$$w_1(t) = \sum_{i=1}^{t} p(i) \quad (4)$$

$$w_2(t) = \sum_{i=t+1}^{L} p(i) = 1 - w_1(t) \quad (5)$$

L is as number of gray level. Then mean of object and background is calculated by using Equation (6) and (7).

$$m_1(t) = \sum_{i=1}^{t} i.p(i) / w_1(t) \quad (6)$$

$$m_2(t) = \sum_{i=1}^{t} i.p(i) / w_2(t) \quad (7)$$

Variance is calculated by following equation.

$$\sigma^2(t) = \sigma_w{}^2(t) + \sigma_B{}^2(t) \quad (8)$$

Opening operation can repair object contour with eliminated pixel area that smaller than structure element [19]. Opening operation can define as following.

$$AoB = (A\Theta B) \oplus B \tag{9}$$

Feature extraction of segmented optic disc is conducted to know significant features that can use for identifying optic disc area. In this work, features are obtained based on texture, color and form feature.

Texture feature is performed by using texture calculation of second order with considered relation of pixel adjacent. It is called gray level co-occurrence matrices (GLCM). GLCM used four direction namely 0º, 45º, 90º and 135º. Features of GLCM are angular second moment (ASM), contrast, inverse different moment (IDM), entropy and correlation. Mathematical formulated of these features are described in following equation.

$$ASM = \sum_{i=1}^{L}\sum_{j=1}^{L} GLCM(i,j)^2 \tag{10}$$

$$contrast = \sum_{n=1}^{L} n^2 \left\{ \sum_{|i-j|=n} GLCM(i,j) \right\} \tag{11}$$

$$IDM = \sum_{i=1}^{L}\sum_{j=1}^{L} \frac{GLCM(i,j)}{1+(i-j)^2} \tag{12}$$

$$entropy = -\sum_{i=1}^{L}\sum_{j=1}^{L} GLCM(i,j)\log(GLCM(i,j)) \tag{13}$$

$$correlation = \frac{\sum_{i=1}^{L}\sum_{j=1}^{L}(ij)(GLCM(i,j)) - \mu_i'\mu_j'}{\sigma_i'\sigma_2'} \tag{14}$$

Colors feature are obtained from statistic calculation consist of mean, deviation, skewness and kurtosis. These features are mathematically formulated respectively in following equations.

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij} \tag{15}$$

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(P_{ij}-\mu)^2} \tag{16}$$

$$\theta = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(p_{ij}-\mu)^3}{MN\sigma^3} \tag{17}$$

$$\gamma = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(p_{ij}-\mu)^4}{MN\sigma^4} - 3 \tag{18}$$

Form feature is obtained with calculating wide and diameter of optic disc area. This result can use for obtaining roundness (R) of optic disc area [19]. Mathematical formulation of R is described in Equation 19.

$$R = 4\pi \frac{A(R)}{P^2(R)} \tag{19}$$

3 Result and Discussion

In this works 39 images are selected from DRISHTI-GS retinal images dataset with PNG format [10]. Selection based on quality images. Red channel as the result of extracted image showed that contrast between of optic disc and background is better than others. It is represented in Figure 3.

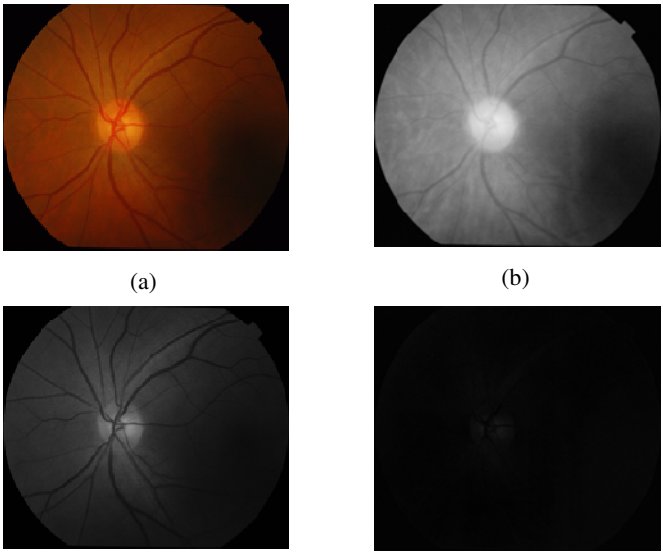
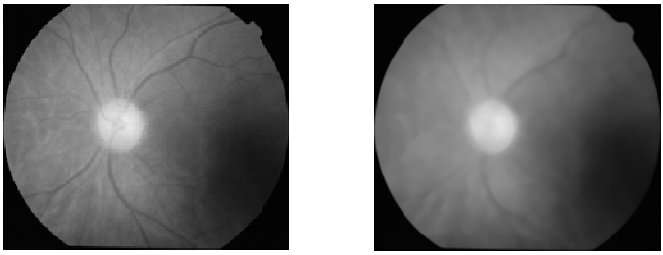


Fig. 3. The result of the extracted channel image. (a) Original image. (b) Red channel. (c) Green channel. (d) Blue channel.

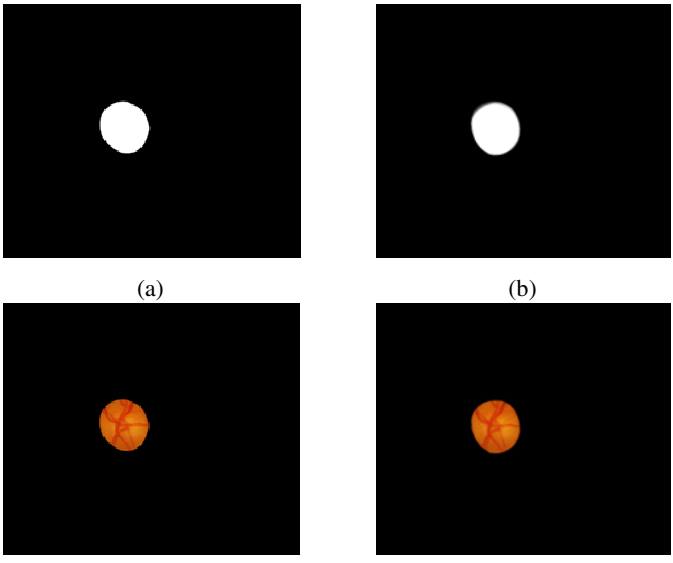
Median filtering is applied to the red channel to remove noises and make optic disc area more clearly. Also, it can reduce visible of blood vessel in optic disc area. Figure 4 shows the result of the filtering processed.



(a) (b)

Fig. 4. The result of filtering processed. (a) Red channel image. (b) Filtered image.

Threshold value that is obtained from Otsu method is very affected to separate optic disc area from background and others object. Segmented threshold image is processed by using opening operation to obtain better segmented optic disc area. This process obtained binary image. Furthermore, segmented imaged is validated with ground truth image to know accuracy of this process. Segmented image is shown in Figure 5.



(a) (b)

(c) (d)

Fig. 5. Segmented image. (a) Segmented binary image. (b) Ground truth. (c) Segmented optic disc. (d) Segmented ground truth.

To obtain accuracy of this work, we calculated wide of optic disc area in segmented images. Result of calculation is compared to wide of ground truth images. Result showed that accuracy achieved 94.546%.

Images dataset are also tested with other technique based on fuzzy C-means (FCM) and morphology. Red channel from original image is extracted. Segmented optic disc that obtained by using FCM and morphology is presented in Figure 6. Accuracy of this performance technique is 68.86%.

Fig. 6. Segmented image by using FCM and morphology

Diameter of segmented images is calculated to obtain center point of optic disc area. It is performed by finding average from totaling between point (x1, y1) and point(x1, y2) from diameter. Figure 7 shows centre point of optic disc. The final result of optic disc localisation is depicted in Figure 8.

Fig. 7. Center point of optic disc area

Blue line circle is optic disc localisation by using proposed method while yellow line circle is ground truth localisation. Feature extraction is conducted by using GLCM as feature texture, color histogram as feature color, and roundness ratio of segmented optic disc that is obtained from wide object. Results show that average of roundness ration from segmented optic this is 0.864. Three features that very significant from red channel of this work are contrast for texture feature, mean and deviation for color feature. Average of these three features are 11.155, 6.859 and 34.406, respectively. Table 1 and table 2 show comparison of average features on red, green and blue channel.

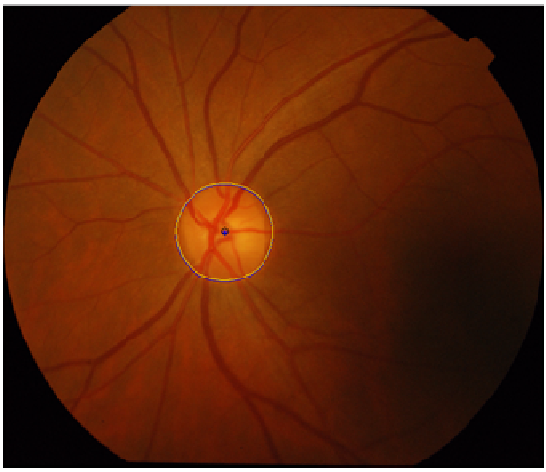


Fig. 8. Optic disc localisation

Table 1. Averages texture features

Channel	Asm	Contrast	Idm	Entropy	Correlation
Red	0.928	**11.155**	0.982	0.361	0.001
Green	0.928	2.708	0.980	0.391	0.004
Blue	0.931	0.439	0.981	0.344	0.131
Grayscale	0.928	3.967	0.981	0.377	0.003

Table 2. Averages color features

Channel	Mean	Deviation	Skewness	Kurtosis
Red	**6.859**	**34.406**	5.505	29.859
Green	3.380	17.835	6.163	40.106
Blue	1.022	5.727	6.927	53.353

Contrast is used to measure intensity differentiation of images. Red channel resulted highest contrast, indicated that red channel has high variation of intensity than others. Between object optic disc and background can separate clearly on red channel. And also, mean and deviation of color features. Mean is average of intensity while deviation is average of contrast. These features can use as parameter to identify optic disc area on red channel than other channels.

4 Conclusion

A method to detect optic disc has been developed. The method consists of two main steps, i.e. preprocessing and segmentation. Median filter is proposed in the preprocessing to remove noises from images. The output of the preprocessing is an image

with better contrast of optic disc. The optic disc is segmented based on Otsu method and morphological operation. The segmented image is validated with the ground truth image. This research work achieves an accuracy of 94.546%. Three features extracted from the optic disc area are contrast, mean and deviation.

References

1. Pachiyappan, A., Das, U., Murthy, T., Tatavarti, R.: Automated diagnosis of diabetic retinopathy and glaucoma using fundus and images. Lipids Health Dis. 11(1), 1–10 (2012)
2. Tjandrasa, H., Wijayanti, A., Suciati, N.: Optic nerve Head Segmentation Using Hough Transform and Active Contours. Telkomnika 10(3), 531–536 (2012)
3. Paranjpe, M.J., Kakatkar, M.N.: Review of Methods for Diabetic Retinopathy Detection and Severity Classification. Int. J. Res. Eng. Technol. 03(03), 619–624 (2014)
4. Kavitha, K., Malathi, M.: Optic Disc and Optic Cup Segmentation for Glaucoma Classification. Int. J. Adv. Res. Comput. Sci. Technol. 2(1), 87–90 (2014)
5. Ning, D., Yafen, L.: Automated identification of diabetic retinopathy stages using support vector machine. presented at the 2013 32nd Chinese Control Conference (CCC), pp. 3882–3886 (2013)
6. Reza, A.W., Eswaran, C., Dimyati, K.: Diagnosis of Diabetic Retinopathy: Automatic Extraction of Optic Disc and Exudates from Retinal Images using Marker-controlled Watershed Transformation. J. Med. Syst. 35, 1491–1501 (2010)
7. Ponnaiah, G.F.M., Baboo, C.D.S.S.: Automatic optic disc detection and removal of false exudates for improving retinopathy classification accuracy. Int. J. Sci. Res. Publ. 5(3), 1–7 (2013)
8. Aquino, A., Gegúndez-Arias, M.E., Marín, D.: Detecting the Optic Disc Boundary in Digital Fundus Images Using Morphological, Edge Detection, and Feature Extraction Techniques (2010)
9. Godse, D.A., Bormane, D.S.: Automated Localization of Optic Disc in Retinal Images. Int. J. Adv. Comput. Sci. Appl. 4, 65–71 (2013)
10. Download Citra, Drishti (March 12, 2014), `http://cvit.iiit.ac.in/projects/mip/drishti-gs/mip-dataset2/Submit-results.php` (accessed: March 12, 2014)
11. Hani, A.F.M., Nugroho, H., Nugroho, H.A., Izhar, L.I., Ngah, N.F., George, T.M., Ismail, M., Hussein, E., Pin, G.P.: Toward a fully automated DR grading system, vol. 37, pp. 663–666 (2011)
12. Rashid, S., Shagufta: Computerized Exudate Detection in Fundus Images Using Statistical Feature based Fuzzy C-mean Clustering. Int. J. Comput. Digit. Syst. 2(3), 135–145 (2013)
13. Eadgahi, M.G.F., Pourreza, H.: Localization of hard exudates in retinal fundus image by mathematical morphology operations, pp. 185–189 (2012)
14. Abbadi, N.K.E., Saadi, E.H.A.: Automatic Detection of Exudates in Retinal Images. Int. J. Comput. Sci. Issues 10(2), 237–242 (2013)
15. Ranamuka, N.G., Meegama, R.G.N.: Detection of hard exudates from diabetic retinopathy images using fuzzy logic. IET Image Process. 7(2), 121–130 (2013)
16. Fang, G., Yang, N., Lu, H., Li, K.: Automatic Segmentation of Hard Exudates in Fundus Images Based on Boosted Soft Segmentation. presented at the International Conference on Intelligent Control and Information Processing, Dalian, China (2010)

17. Yazid, H., Arof, H., Mohd Isa, H.: Exudates segmentation using inverse surface adaptive thresholding. Measurement 45(6), 1599–1608 (2012)
18. Prasetyo, E.: Pengolahan Citra Digital dan Aplikasi dengan Matlab. ANDI, Yogyakarta (2011)
19. Kadir, A., Susanto, A.: Pengolahan Citra Teori dan Aplikasi. ANDI, Yogyakarta (2012)
20. Putra, D.: Pengolahan Citra Digital, I. ANDI, Yogyakarta (2009)

Visualization Techniques

Multimedia Design for Learning Media of Majapahit

Silvia Rostianingsih, Michael Chang, and Liliana

Informatics Departments, Petra Christian University
Siwalankerto 121-131 Surabaya, Indonesia
{silvia,lilian}@petra.ac.id, m26411052@john.petra.ac.id

Abstract. Majapahit was one of the last major empires of the region and is considered to be one of the greatest and most powerful empires in the history of Indonesia and Southeast Asia. However, learning history is became unpopular to young generation. Interactive media can help society to learn history enjoyable. This interactive multimedia will allow the reader to interact with combination of text, image, sound, and animations of Majapahit. There are three aspects of disccusion which are economics, politics, also arts and cultures. Each aspects will provide information related with Majapahit. From the results of questionnaire can be conclude that interactive multimedia form is more interesting than history's books which full of text. The audience is satisfy with the utility and interaction form. Their knowledge also improve after they use the application.

Keywords: Majapahit, multimedia, interactive.

1 Background

Majapahit was one of the last major empires of the region and is considered to be one of the greatest and most powerful empires in the history of Indonesia and Southeast Asia. Majapahit was a vast archipelagic empire based on the island of Java from 1293 to around 1500. However, learning history is become unpopular for young generation. Interactive media can help society to learn history enjoyable. Interactive learning media can be use as a learning tool individually (Siagian, 2014). Multimedia technology enhance the knowledge (Kaewkiriya, 2013). Moreover, the interactive media can be accessed through devices that support the technology to open this interactive media. In order to improve the egarness of learning history, this research design a multimedia for learning media of Majapahit. This application is built by Unity 3D which is easily handle complex animation.

2 Majapahit

Majapahit was founded by Raden Wijaya. It started when Jayakatwang attacked Singasari and causing Kertanegara killed. At the time when Java Mongol attack Jayakatwang, Raden Wijaya pretend to help the Mongols. After Jayakatwang killed,

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 363–371, 2015.
DOI: 10.1007/978-3-662-46742-8_33

Raden Wijaya turns against the Mongols and managed to repel the Mongol armies from the land of Java. In 1293, Raden Wijaya ascended the throne and the title of Sri Kertajasa Jayawardhana.

There are 4 (four) aspect to be concluded in the design: (1) economics, (2) politics, and (3) arts and culture.

2.1 Economics

Majapahit was an agrarian and maritime empire. Foreign traders who came to Majapahit came from Campa, Khmer, Thailand, Burma, Sri Lanka, and India. Majapahit is also cooperating with several countries, such as, Campa, Cambodia, Siam, Burma, and Vietnam southern (Aji, 2014). Commodities are brought to the Majapahit are silk and china ceramics, fabrics from India, and incense from Arabic. The goods were exchanged for spices and other agricultural products. Ceramics are came from China and Vietnam (Annam) in the form of plates, bowls, and large glasses. While from Thailand, Majapahit import plates, bowls, and small glasses.

2.2 Politics

A king was the highest authority in Majapahit, assisted by Rakryan Mahamantri Kartini occupied by the son of the king and the governing board. They are consisting of Rakryan Mapatih or Patih Mangkubumi, Rakryan Tumenggung, Rakryan Demung, Rakryan Rangga, and Rakryan Kanuruhan. This structure is located in the central government and to areas that are under the king has a similar structure. Majapahit also established relations with other countries to expand trade and to the expansion of the territory of Majapahit.

2.3 Arts and Cultures

Majapahit's arts were temples and literaries. There are many temples such Penataran, Sawentar, Tigawangi, Surawana, and Rat. Some temples contained narrative relics and some only had plain temple without relics. The relics are contains history that occurred at that time. For examples, Surawana's relics informs the daily life from Majapahit. The most famous literary works in the Majapahit period which is the book of Negarakartagama which is written by Prapanca (Gunawan, 2013). Art at the Majapahit lies not only in temples and literaries, but there are also dance and puppets. At the time of Majapahit, arts already exists. However, during the reign of the Majapahit arts gradually weakened. Arts are experiencing growth after the collapse of Majapahit post (Aji, 2014).

3 Multimedia

Multimedia uses different media to convey information as text, audio, graphics, animation, and video, all done interactively (Catwright, 2006). Digital multimedia is a

combination of two or more media which is represented in digital form, well integrated to be presented, or manipulated by computer programs (Chapman, 2000). Multimedia is divided into 5 format which are:

1. Multimedia as a Tutorial
This format is an interactive multimedia learning tutorials conducted like an instructor or teacher. This format presents the information in the form of text, images, either still or moving, video, sound and graphics.

2. Multimedia as Drill and Practice
This format is used to train skills so the users have a proficiency in a skill or mastery strengthens the theory of a concept. This format is also equipped with the correct answer and an explanation of the answer.

3. Multimedia as Simulation
This format has the learning objectives which are trying to emulate a process that occurs in the real world so that users can experience the real world simulations. Examples aircraft simulation.

4. Multimedia as an experiment and Experiment
This format is almost the same as the simulation only this format is usually found in the laboratory of biology, physics, and chemistry that has practical activities so the user can develop experiments.

5. Multimedia as Games
Games are intended to archieve learning process by playing game.

The interactive multimedia elements are text, picture, sound, animation, video. The text is the basic of the information provided by application. The text itself is composed of several symbols, alphabet, and numbers. Image has always been an important part in human oriented multimedia because of the visual (based on sight). Voice is one of the multimedia elements which is interesting and unique than the other multimedia elements. Sound provide information to users and indulge the senses of hearing from user. Animation is a sequence of static images. Animation usually made specifically to help users to illustrate a concept that is becoming a learning requires animation. A video shows the picture of life, so the information is more accurate and interesting.

In order to create interactive multimedia, we must determine objectives of the project. The purpose of the project maybe seen already very obvious but here meant more specific and clear as possible to assist within the project. Second, we must understand the audience, include age range, sex, particular interest, and also level of education. Notice also what the audience experienced a hearing or visual impairments so on. Through the multimedia, audience will be passive or active and also educate or simply entertain. Third, design information should be neatly arranged in a group and not jumping from one to the other information. Forth, interface design should be clear, structured, and understandable. Last, we must test the result to make sure each finished work on a project. Well-designed materials induced positive emotions and facilitated comprehension (Plass, 2014).

4 Multimedia Design

4.1 Structure of Content

The media is divided into 4 (four) aspects which are (1) History of Majapahit, (2) Economics, (3) Politics, and (4) Art and Culture. The economics explains trade's lane and rivers as a major role in the trade at the time. It also explains which countries are liaised with Majapahit. The politics explains the kings whom rule the kingdom. The arts and cultures explains the historical relics. The structure would be:

1. Majapahit
   (a) History
   (b) Genealogy of the kings
   (c) Rebellion
   (d) Expansion
2. Economic
   (a) Agriculture
      (i) Type of farming
      (ii) Farming technology
   (b) Trade
      (i) International trades
      (ii) Goods
      (iii) Currency
      (iv) Market
   (c) Industry
      (i) Type of industry
      (ii) Technology
   (d) Tax
      (i) Description
      (ii) Type of tax
      (iii) Tax collection mechanism
3. Politic
   (a) Society and the political system
      (i) System
      (ii) Environment
4. Art and culture
   (a) Ancient
      (i) Temples
   (b) Religion
      (i) Siwa-Budha
      (ii) Islam
   (c) Revolution
   (d) Ceramic craft
   (e) Statue
   (f) Metal
   (g) Performing
   (h) Literature

4.2 Multimedia

Multimedia template is divided into 6 (six) which are image, animation, 3D object, initial menu, menu and blank section. Button is design with a vintage style considering the year of Majapahit. Background is relating to the history of the paper that has a very long and brownish in color so it seems like in historical times. Shapes both alone and in conjunction with warm color induced positive emotions (Plass, 2014).

3D object templates display explanations accompanied by a 3D object and description. User can rotate the 3D object by the y-axis and the x-axis. Image template is a template that contains explanations with images relating to the material being described in the template. Animation template associated with an explanation of the content of the material. The purpose of the animation is to help users understand the content of the material being described in the template. Blank template is a template used to display both images, animation, video, 3D objects, as well as the content of the material. The font's type for the description or picture's title are Linotypes Palatino with 15pt size. The font's for the chapter's title is Algerian with 18pt size. The font's type for each button is using Forte with 18pt size.

Table 1. Template

Title	Template
Trade	Image
Goods	3D Object
Market of trade	Text
Tax	Image
Tax collection mechanism	Animation
Tax allocation	Text
Religion	Image
Politic	Image
Art and culture	Image

5 Implementation

In the main district of Majapahit, user can view the location by click one of the 12 district. It is also give information who is ruled and the relation of the king. There are 12 buttons with animated color. The color divided into three status: normal (light color), on-hover (normal color), pressed (dark color).

International trade animate the trade lanes by display an object which is move from one place to another place. The object will remain elliptical oval if located on the trade route passing through the sea. The object will change into a round shape and when the trade lanes located on land.

In the culture menu, there is a button marked "king". When the button is pressed, it will display the king who ruled and it will enlarge the picture of king. Pictures of the king who ruled the kingdom of Majapahit is a set of buttons that given the background image of the kings who ruled the kingdom of Majapahit. Animation is given on each button which provide animation in position, size, rotation, color, and

magnification. When users make a hover, the image will be enlarged in size to 400x400 and the position of the image is also changed to be in the right side of the page. If the user has not click the button hover, the image automatically shrink according to the size of the button.

Hampir disetiap halaman dapat dijumpai kotak kecil berisikan tanda Tanya yang berputar 360 derajat pada sumbu Y. Kotak tanda tanya tersebut berisi informasi tambahan mengenai penjelasan yang dibahas pada halaman tesebut yang terletak di belakang kalimat yang ingin diberi informasi tambahan dengan mengeluarkan kotak yang berisi infomasi tambahan tersebut ketika user melakukan hover pada kotak tanda tanya tersebut. Cara penganimasiannya menggunakan fitur animasi dari button, kotak tanda tanya yang berputar 360 derajat menggunakan component button script animasi berupa animation yang di-auto-generate sehingga bebas memberikan animasi dalam posisi, ukuran, rotasi, warna maupun perbesaran. Dalam keadaan normal kotak tersebut diberikan animasi per 30 detik berputar sebesar 90 derajat sampai pada detik ke-2 menjadi 360. Animasi dari button diulang secara otomatis sampai user berpindah halaman ke halaman selanjutnya. Ketika user melakukan highlighted pada button maka dilakukan animasi pada "caption" yang merupakan child dari parent button dengan mengubah ukuran, posisi dan warna pada "caption" yang berisi informasi tambahan dari penjelasan materi yang ditempatkan pada posisi tertentu di halaman tersebut. Apabila user sudah tidak melakukan hover pada button maka "caption" secara otomatis ukurannya kembali seukuran dengan button dan posisi dari "caption" juga kembali pada posisi dimana letak button berada.

Almost in every page can be found a small box containing question marks that rotates 360 degrees on the Y-axis. The question mark box contains additional information when the user hover on the question mark box. The box that rotates 360 degrees is using the component button animation script that are auto-generated so freely give animation in position, size, rotation, color and magnification. Under normal circumstances the box is given a 30 second animation rotates by 90 degrees until the second-to-2 to 360.

6 Results

The sub menu of economics section is the origin of goods. Fig. 1 shows the globe spin to match the position of the selected country (Khmer country).

In order to describe the 12 district of Majapahit, user can choose the district's button and the map will show the area with a bright color and brief information about the area (name and image of the king). Fig. 2 shows the information of Matahun's districts which rules by Rajasawardhana. Fig. 3 shows the animation of the trade's lane. The green cirle will move along the lane path to indicate the route.

This learning media is tested againts 32 people. The results of questionnaire can be seen in Table 2. The results are 96,9% found the utility of the application is good, 78,1% found the interactive of the application is good, 62,5% fount that their knowledge is improve, and 56,3% are eager to learn more about Majapahit.

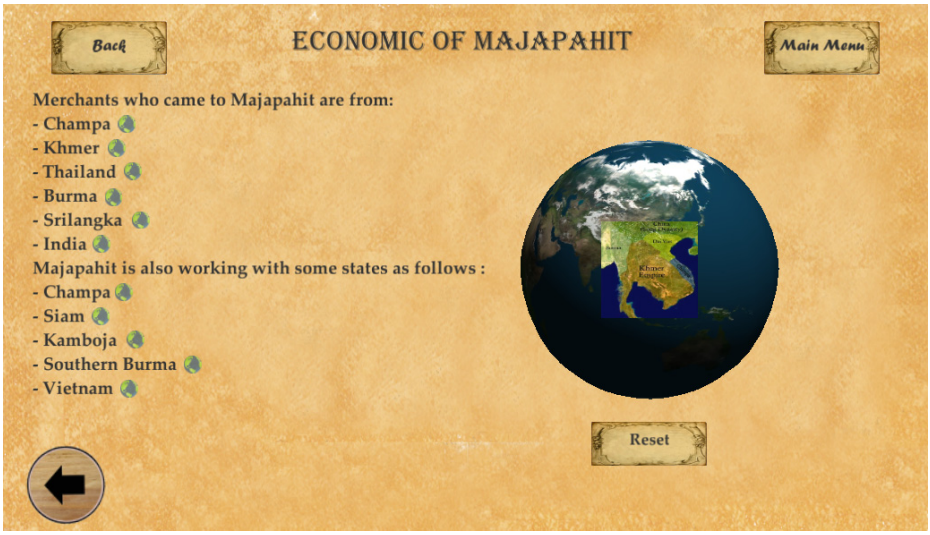


Fig. 1. Origin of Goods

Fig. 2. 12 Main Disctricts

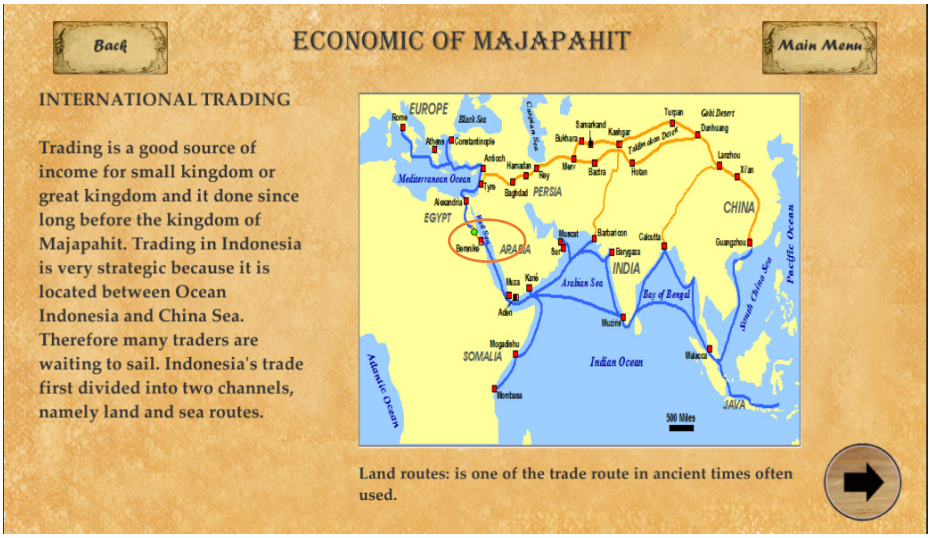


Fig. 3. Lane of Trade

Table 2. Results of Questionnaire

Description	Good	Average	Poor
Utility	96,9%	3,1%	0,0%
Interactive	78,1%	15,6%	6,3%
Improve your knowledge	62,5%	37,5%	0,0%
Willingness to learn more	56,3%	28,1%	15,6%

7 Conclusions

From the results of questionnaire can be conclude that interactive multimedia form is more interesting than history's books which full of text. The audience is satisfy with the utility and interaction form. Their knowledge also improve after they use the application.

Acknowledgment. We would like to thank the KOPERTIS disctrict VII Jawa Timur for the support of this work through project (DIPA Kopertis No. 023.04.2.415015/2014 dated 5 Desember 2013).

References

1. Plass, J.L., et al.: Emotional design in multimedia learning: effects of shape and color on affect and learning. Learning and Instruction (2013), http://dx.doi.org/10.1016/j.learninstruc.2013.02.006

2. Aji, K.B., Achmad, S.W.: Ensiklopedia raja-raja nusantara: menyingkap tuntas riwayat hidup raja-raja nusantara. Araska, Indonesia (2014)
3. Chapman, N., Chapman, J.: Digital cartography, 3rd edn. John Wiley & Sons (2009)
4. Gunawan, R.: dkk: Sejarah Indonesia kelas X. Politeknik Negeri Media Kreatif, Jakarta (2013)
5. Siagian, S., Mursid, Wau, Y.: Development of interactive multimedia learning in learning instructional design. Journal of Education and Practice 5(32) (2014)
6. Kaewkiriya, T.: A design and development of e-learning content for multimedia technology using multimedia game. International Journal of Software Engineering and Applications (IJSEA) 4(6) (2013)
7. Cartwright, W., Peterson, M.P.: Multimedia cartography, 2nd edn. Springer (2007)

Adding a Transparent Object on Image

Liliana, Meliana Luwuk, and Djoni Haryadi Setiabudi

Informatics Department, Petra Christian University, Surabaya, Indonesia
{lilian,Djonihs}@petra.ac.id, m26409027@john.petra.ac.id

Abstract. Nowadays, image manipulation, which is related to matting and compositing process, is commonly used. Processed object may vary, however matters may arise when the object is transparent. Not many research has been conducted about the issue due to the difficulties in extracting transparency value and object refraction. In this research an application to execute matting process of a transparent object and compositing it into a new image is developed.

Matting process will be executed using Grabcut method, followed by calculation of transparency value of an object using alpha matting. The transparency value is then used in compositing an object with a new background, while refraction will be executed using refraction calculation of ray tracing. Adding a transparent object into new image (new background) process, first user should determine the eye's position and distant between new image and transparent object. Then, assume the additional transparent object as a screen to trace rays from eye. Each ray will be bent and hit the new image. The color of new image will add into the transparent object area.

Experiments result showed that extracted transparency value was affected by lighting, refraction of the old background, which unknown and object's refraction index. Refraction result looks natural compared with the actual condition. Moreover, eye position can be configured by the user to obtain results as desired.

Keywords: Matting, Compositing, grabcut, Transparency, Refraction, Alpha channel.

1 Background

The need of manipulating an image is common in many aspects, especially in advertising. One of common image manipulating applications is matting and compositing. That are cropping a distinct object from an image and adding the cropped object into a new image as a new background. A matter in doing matting and compositing is if a transparent object is involved. A transparent object shows other object behind it. The color of background object will be blended with the color of the transparent object. It is impossible to separate the blended color when the transparent object is taken from an image. It is difficult to determine which is background's color or the transparent object's color.

In research done by Gong, Liang, Yang R., and Yang, Y.H. in 2010, alpha channel with dot product is used [1]. It separates the interpolation color between foreground

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 372–381, 2015.
DOI: 10.1007/978-3-662-46742-8_34

and background. This alpha channel is computed based on the statistical color from trimap. Trimap is given area of foreground area, background area and unknown area between those two other areas [2,3,4,5]. Tomas Malmer proposed the combination of statistical calculation and grabcut [6,7] to separate pixel in unknown area which are that belong to foreground, which ones belong to background [8]. Zongker and Han et.al improved the segmentation method over natural environmental or textured background [7,9].

In our research, we develop an application to matte a transparent object using alpha channel and grabcut, and composite it into a new image using refraction mechanism of ray tracing.

2 Refraction Mechanism of Ray Tracing

A ray will be traced through the air. When it hit a transparent object, it will be refracted or reflected. The refracted or reflected ray will be traced again[11,12]. This process will be done until it hits a non-transparent object. We modify this mechanism to fit with our case. First the new image will be the non-transparent object. The adding object is assumed as a grid screen where rays will be traced. This grid screen will be placed some distant from the background. Then, an eye will be positioned in front of them. New image, transparent adding object and the eye arrangement like seen on fig.1

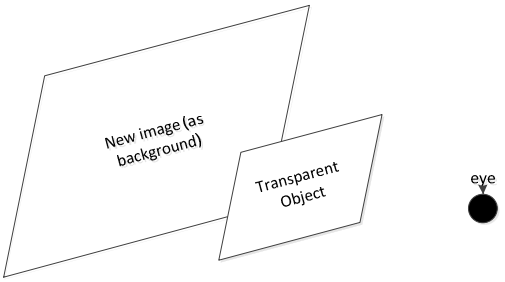


Fig. 1. The position of new image, adding object and eye

Every transparent object has a refraction index. It determines how far the ray will be refracted. Smaller the refraction index further the refracted ray will be. The refracted ray will be computed using six steps shown in fig 2 [11]. The important variable in computing the refraction ray are angle of coming ray, θi (from the air into the transparent object) and angle of the refracted ray, θt (from the transparent object outside). The angle of coming ray is computed using dot product between the

ray from eye to the screen of transparent object and the normal plane of the screen. This dot product can be seen in eq. 1 and the refracted ray can be counted using Snell's law as shown in eq. 2

$$\theta i = \cos^{-1} \frac{(eye\ position - pixel\ position) \cdot (normal\ plane)}{|(eye\ position - pixel\ position)||(normal\ plane)|} \tag{1}$$

$$\frac{\sin \theta i}{\sin \theta t} = \frac{n2}{n1} \tag{2}$$

in eq. 2, n1 is refracted index of the air (1.0003), n2 is refracted index of the transparent object.

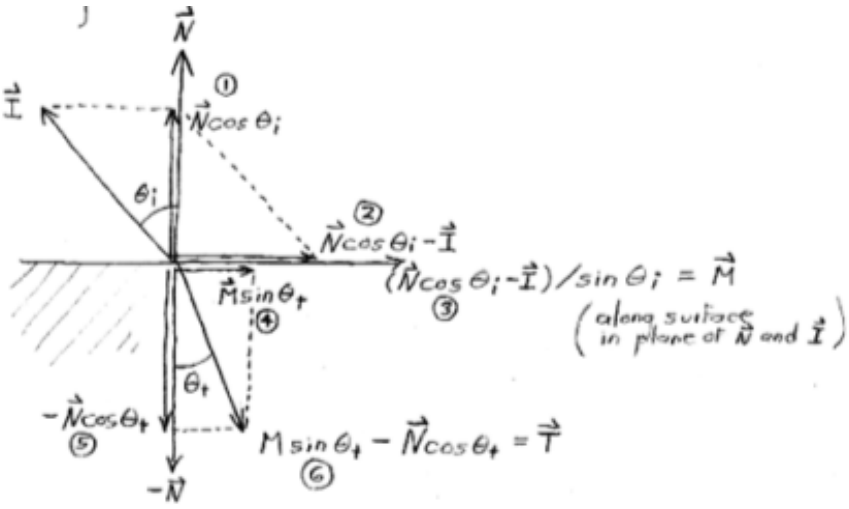


Fig. 2. Six steps to compute the refracted ray[11]

After compute the refracted ray, the refracted ray will be trace until hit the new image behind it. The color of the pixel which is hit by the refracted ray will be added into the color of the pixel on the transparent object where the ray comes from. To find the position of the pixel which is hit, we use intersection between a line and a plane, shown in eq.4 First, the time when the refracted ray hit the new image is count. Then, the time will be used in the eq. 3 [11, 12].

$$t = \frac{d - N \cdot S}{N \cdot refracted\ ray} \tag{3}$$

$$pixel\ position\ which\ is\ hit = S + t \cdot refracted\ ray \quad (4)$$

t is time when the refracted ray hit the new image. d is the distant of the new image. in this research, d=0. N is the normal of the new image (0, 0, 1). S is the position of the pixel on the transparent object where the refracted ray comes from. Operation between N and S is a dot product.

3 Object Adding Transparent into a New Background

There are three main processes in this application, define the trimap, crop the target object, and adding the transparent object into a new image. Adding the transparent object is done by interpolating the color of the transparent object with the background color found using eq. 4.

User defines the trimap. After calculate the alpha channel from the unkown area, then the transparent object will be added into a new image as a new background. The diagram that shows the whole process can be seen in fig. 3. The focus of this research is adding the transparent object into a new image.

Fig. 3. The diagram of the whole process from matting until compositing the transparent object

The flowchart that represents the adding transparent object into a new image algorithm can be seen in fig. 4. First, the eye position, the distant between eye and the transparent object and the distant between transparent object to the new image should be given. Ray tracing mechanism needs 3D environment. Thus, the given distant input give the depth for the 2D image. Assuming the transparent object as a screen, the rays from eye trace through the each pixel on the screen. After that, the rays will be bend using eq. 2. The pixel posisition on the screen (row and colom represent x and y axis) will be the start point of the refracted ray. The refracted ray will be traced and checked the intersection with the plane behind it, that is the new image. The color of the hit pixel on the new image will be interpolated with the color of the transparent object multiply with the alpha value. The new color will be the color of the transparent object after combine with the new background.

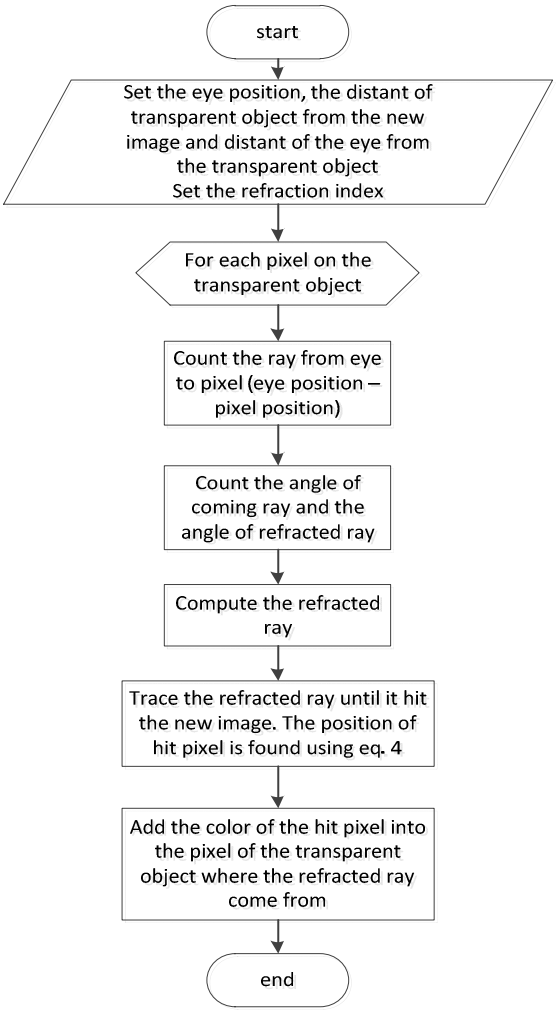


Fig. 4. Flowchart of compositing a transparent object into a new background

4 Experiment

Experiment is done in several aspects, comparison with actual condition, the matting and compositing result, and experiment using two different distant between transparent object and eye. The result of comparison between actual condition with our application result can be seen on table 1, while the matting a transparent object and compositing it into a new image can be seen on table 2. The table 3 shows the result of experiment with “near” and “far” distant of eye position from the transparent object.

Table 1. The comparison between the result of this application with the actual condition

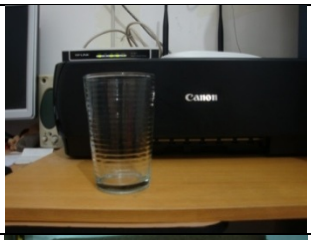
Actual condition	This application's result
	
	
	
	

Table 1 shows similarity between the actual condition and the simulation result of our mechanism. The difficulty in reconstructing the result is to determine the distant between eye and the transparent object and how many refraction index of the material. In this experiment, we use common home's utility so the refraction index can be determined easily.

In order to examine the result yielded by various distant between eye and the transparent object, and vertical position, we set two different positions categorized near and far. The result can be seen in table 2.

Table 2. The result of grabbed object (middle colomn) into a new image (right colomn)

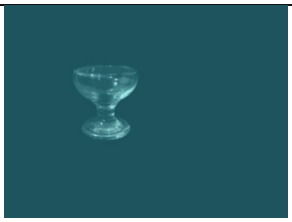
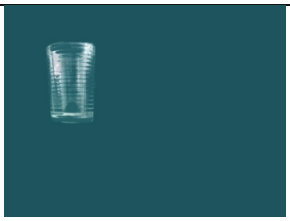
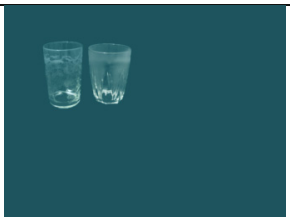
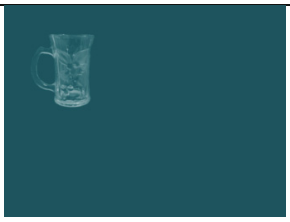
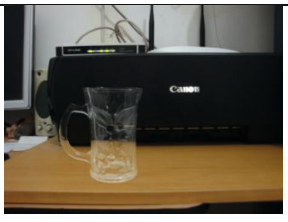
The original image	Cropped transparent object	Compositing into the new image
		
		
		
		
		

Table 2. (*Continued*)

The original image	Cropped transparent object	Compositing into the new image
		
		

Table 3. Result of various eye position toward the transparent object

Near	Far in front of the transparent object
	
	

Table 3. (*Continued*)

Near	Far in front of the transparent object
	

Table 3 shows the appearance of other objects on the transparent object is effected by the eye's position. Further the eye's position, smaller the objects behind it then it will be looked wider the field of view through the transparent object. Further the eye's position, the more distortion of the object behind it will be looked.

5 Conclusion

From the experiment have done, ray tracing mechanism can be implement well in simulating the compositing a transparent object into a new background. The hit position of the new background should be discrete (pixel position) while the computation gives a decimal number. However, this problem doesn't effect the result's appearance.

References

1. Gong, M., Wang, L., Yang, R., Yang, Y.-H.: Real-Time Video Matting using Multichannel Poisson Equations. In: Proceedings of Graphics Interface 2010, pp. 89–96. Canadian Information Processing Society, Toronto (2010)
2. Porter, T., Duff, T.: Compositing Digital Images. Computer Graphics 18(3), 253–259 (1984)
3. Sun, J., Jia, J., Tang, C.-K., Shum, H.Y.: Poisson Matting. ACM Transactions on Graphics – Proceedings of ACM SIGGRAPH 2004 23(4), 315–321 (2004)
4. Wang, J., Cohen, M.: Image and Video Matting: A Survey. Foundations and Trends in Computer Graphics and Vision 3(2), 97–175 (2007)
5. Yeung, S.-K., Tang, C.-K., Brown, M.S., Kang, S.B.: Matting and Compositing of Transparent and Refractive Objects. ACM Trans. Graph. 30(1), Article 2 (2011)
6. Rother, C., Kolmogorov, V., Blake, A.: "GrabCut" – Interactive Foreground Extraction using Iterated Graph Cuts. Proceedings of ACM SIGGRAPH 2004 23(3), 309–314 (2004)
7. Han, S., Tao, W., Wang, D., Tai, X.-C., Wu, X.: Image Segmentation Based on Grabcut Freamework Integrating Multiscale Nonlinear Structure Tensor. IEEE Transactions on Image Processing 18(10), 2289–2302 (2009)
8. Malmer, T.: Image Segmentation using GrabCut (2010)

9. Zongker, D.E., Werner, D.M., Curless, B., Salesin, D.H.: Environtment Matting and Compositing. In: ACM Siggraph 1999, pp. 205–214 (1999)
10. Peng, C.G.: Image Processing: Interpolation (2004)
11. Levoy, M.: Basics of Ray Tracing. Computer Graphics: Image Synthesis Techniques. Lecture notes (1997)
12. Gortler, S.J.: Fondations of 3D Computer Graphics. MIT Press Books, Cambridge (2012)

3D-Building Reconstruction Approach Using Semi-global Matching Classified

Iqbal Rahmadhian Pamungkas and Iping Supriana Suwardi

School of Electrical Engineering and Informatics,
Bandung Institute of Technology,
West Java, Indonesia 40132
iqbal.pamungkas@students.itb.ac.id, iping@informatika.org

Abstract. The complexity problems of a city makes the understanding of three-dimensional spatial structure of urban areas becoming very important, for example in the field of urban planning, telecommunications, updating maps and GIS information, Intelligent Transportation System, and monitoring of urban growth that occurs very rapidly, especially in cities of the developing countries. Techniques for obtaining Digital Surface Model (DSM) of a region continues to evolve, the continued development of technology with high-resolution satellite imagery provides an alternative data and information that is more promising, especially in those countries are not available aerial photography and LIDAR technology. DSM is not sufficient to meet the need for data and information about the city, to generate data that can be understood and utilized furthermore, still needs other processes to be done such as the objects separation process, building with other building, roads, rivers, plants and etc. This paper proposes a model for the 3D reconstruction of buildings automatically using satellite imagery data. In our approach, we propose to use a basic method Semi-Global Matching (SGM) with performed the detection and classification process. This approach aims to increase the accuracy of height estimation, faster computation time and low computational load.

Keywords: 3d reconstruction, building detection, satellite, image, DSM, DEM.

1 Introduction

In the past decade, remote sensing imagery makes monitoring the Earth's surface and atmosphere are possible to take in variety of scales. The improvement of image sensor technology able to produce remote sensing images with better quality. There are many satellite with imaging results such as WorldView and WorldView-2, GeoEye-1, IKONOS, QuickBird, TerraSAR-X, RADARSAT-2 and nanosatellite SkySat-1 which able to produce high-quality video surveillance. Thus, the process and analyzed of the problems from observational data through satellite imagery city become more robust and more effective.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 382–391, 2015.
DOI: 10.1007/978-3-662-46742-8_35

Fig. 1. IKONOS and QuickBird (http://mapmart.com)

So far, data of three-dimensional topography of the land has obtained as form of contour lines or a digital elevation model (DEM) which generated from the processing of satellite data, aerial photographs to measure directly on the ground. Figure 2 below shows the difference in ground surface elevation information. Contour or DEM showed elevation of the ground surface without any objects on top of it whereas, the Digital Surface Model (DSM) covers the objects on the surface.

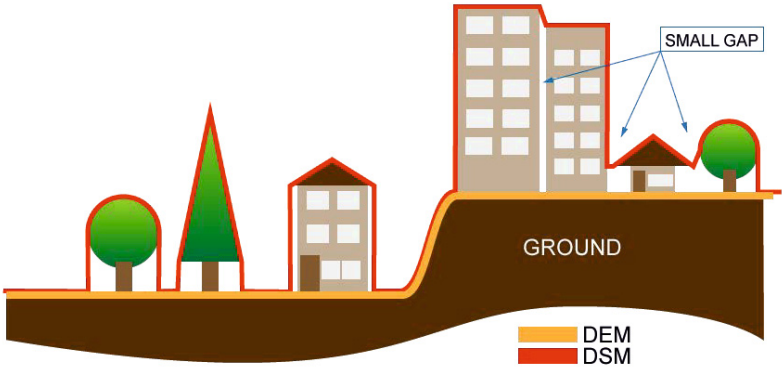


Fig. 2. DEM and DSM

2 Related Research

In order to reveal the detail of the structure of urban space correctly, a single data ground elevation is not enough. Hence, many researchers tried develop the algorithm to obtain the elevation of the roof of the building from various types of data, and the most widely used are aerial photographs and high-resolution satellite image.

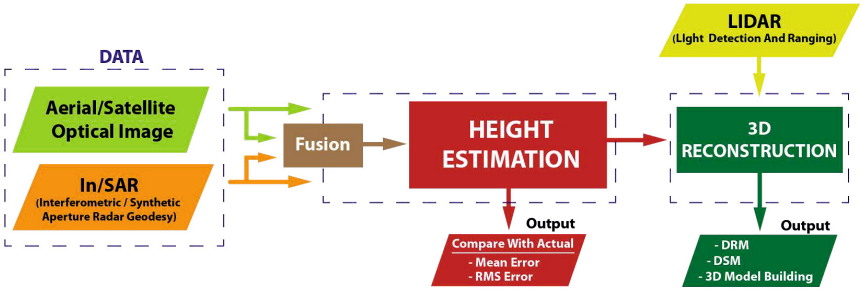


Fig. 3. Previous research mapping

Figure 3 shows a research process diagram of previous studies to determine the height of buildings in urban areas. There are three types of data which used to estimate the height of the building and reconstruct it into a Digital Surface Model (DSM), namely the optical, and radio images from satellites and aircraft, and LIDAR (Light Detection and Ranging).

3 The Problems

The challenges to detect the building and reconstruct it into 3D models are:

- The usage of data There are various types of data. In most cases, many research use stereo or multiple images. Range of data can be generated from the analysis stereo, but the quality is not good enough to generate hypotheses directly because many buildings shortage roof texture sufficient for stereo processing. Moreover, trees which close each other and have same height make data difficult to analyze. The range of accurate data such as LIDAR, are expensive. There is an effort to maximize the use of high-quality data [1, 2] or to improve the coverage quality of image-based data with more than 10 images [3].
- Object extraction we are dealing with satellite images of wide urban areas and many objects are close. It is difficult to separate the boundaries of the building such as boundary lines disrupt roads, plants and other objects around it. Although LIDAR produce point clouds data that can easily be processed into 3D models, it still requires the separation process which is not easy.
- Measurements Height height information obtained from satellite imagery is the height of the building roof or crown of the tree, which different from the conventional model using the Digital Elevation Model (DEM).

- Representation Model representations have an important role in building detection and description of the problem. When we use a simple representation, such as extrusion of the roof forms a rectangle [4, 5], the description will be stronger but the detection rate will be lower because of the limited representational strength. In contrast, when we use representational models is very high, such as polygonal meshes smooth [6, 1], we can describe the building better, but the result will be less strong and the level of geometric information obtained too little, so the usefulness of the results will be very limited. It's important to find a good representation that has the strength and rich of geometric information, which, at the same time, the detection and description algorithm robust and fast computation process.
- High-computing process In order to implement automatically in the area of 400km2 with a resolution of 20cm / pixel [8] (which in comparison, vast cities - big cities in Indonesia, such as Jakarta is 740.3 km2, covering 374,8km2 Surabaya, Medan area 265,1km2, and Bandung area of 167.7 km2). By using the clustering system of 40 pieces Opteron 2.0 GHz CPU and 2.2 GHz, takes 3 to 4 days.
- Semi-automatic method The operator identifies the boundaries of the building in a set of satellite images. The dimension of the building is obtained by using the acquisition geometry, image transfer, and perspective projection. These processes are exhausting and high time consuming. So, automatic procedure to build DSM include the shape of the building based on image matching techniques is needed.

In this research, we used data from satellite optical imagery. So, the process diagram for this research based on the previous researchers as shown in Figure 4.

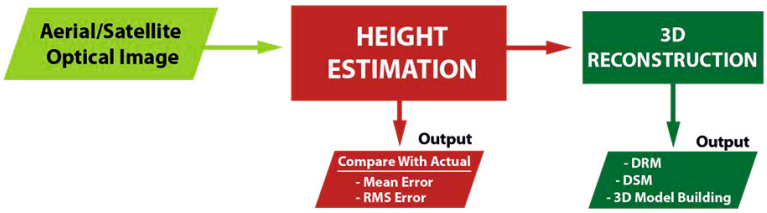


Fig. 4. The research process diagram

There is a research to generate information of the surface of the Earth include the objects on it to DSM which has been conducted since 1982 by Frstner by introducing the Least Squares Matching (LSM) method. In 2003, Hashiba, et al. [7] have focused on the problem of building heights in urban areas. By using a stereo image of IKONOS satellite sensing. Hashiba, et al. [7] introduced the concept of the Digital Roof Model (DRM). DRM provide data on the height of the roof of the building on urban areas which very dense of office buildings and housing.

Figure 5 describing the building roof height measurement principles of satellite positions S1 and S2. I1 and i2 angle shows the angle of incidence of each individual satellite positions, and shows the difference angle in azimuth. One building is taken as a reference altitude, and then DRM automatically built for the building - the surrounding buildings. The result of the research is a 3D Bird's Eye View (Figure 6) from the Digital Roof Model with an average error rate of 5.2 meters.

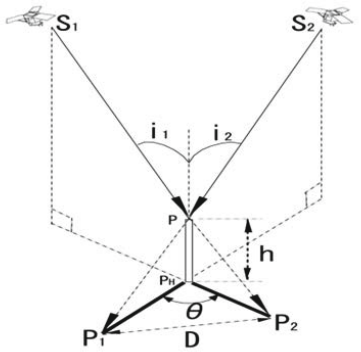


Fig. 5. The principle of measurement of building height over satellite positions S1 and S2 [7]

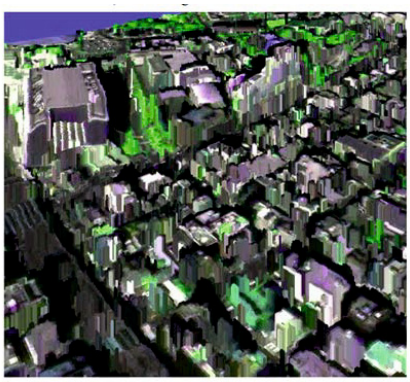


Fig. 6. 3D Birds Eye View in Sumida-ku area [7]

In 2009, Izadi, M, et al. [9], had done some research to present a methodology to estimate the height of the building which has a flat roof and a polygonal shape. The data gathered from a satellite or aircraft imagery. Input from the roof polygonal shape of the building inserted manually. Shadow and shape from the roof of the building used as the main parameter in estimating heights. A fitness function used fuzzy rules for evaluating the estimation of altitude. The real altitudes taken by a genetic algorithm in the search space. The results of

the study are data of the height with average of errors 0.35 meters and root mean - average of 0.96 meters. Izadi, M, et al. [11], had continue his research by increasing the complexity of the shape of the building's roof (polygonal and curved). In 2009, A. Alobeid, et al [10], had done a comparison of methods to estimate the height of the buildings in an urban area using high-resolution stereo photos from the IKONOS satellite. The methods are the Least Squares Matching (LSM) in a region growing fashion by Forstner (1982), Dynamic programming (DP) which referenced in the research Birchfield and Tomasi (1999) and Semi Global Matching (SGM) and Mutual Information (MI) developed by Hirschmller (2008). Compared of three methods, Semi-Global Matching and Mutual Information have the best accuracy.

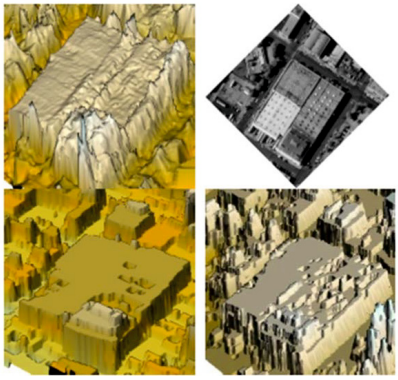


Fig. 7. 3D View from the roof of a building in San Diego [10]

4 Theory

Extracting the information of multispectral images use a pre-process separation the vegetation objects and shadows.

4.1 Linearized Vegetation Indices

The common method used to recognize the vegetation object is normalized difference vegetation index (NDVI). Based on some research, NDVI showed the most effective performance [12]. Measuring the NDVI use this equation [13]:

$$\text{NDVI} = \frac{\text{NIR} - R}{\text{NIR} + R},$$

R and NIR are representing red and near-infrared bands.

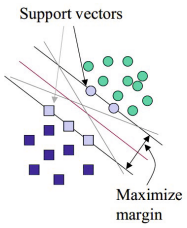
4.2 Linearized Shadow and Water Indices

The other pre-process to recognize the objects are identify the shadows and water, it called shadow-water indices (SWI). The advantages of SWI are recognize the river in a residential area, shadows and cloud which captured by satellite sensors. Shadows also have photometric characteristics similar as water. Representation of shadow and water consists of high value of blue. Ideally, there are green and red bands, but the green band also responds strongly as vegetation and it can disguised shadows or water observation. Hence, the index should include at least the blue and red bands. To obtain the index, the same framework used for the derivation of NDVI with emphasis on component bands of blue, red and near infrared. So it can be written as follows [12]:

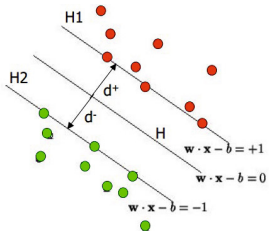
$$\gamma_1 = \frac{4}{\pi} \arctan\left(\frac{\rho_{\text{nir}} - \rho_{\text{blue}}}{\rho_{\text{nir}} + \rho_{\text{blue}}}\right)$$

4.3 Image Classification

In the process of measuring the estimated height of the building, after a preprocessing NDVI and SWI, image classified and formed into a thematic map. There are several methods of classification that can be used for image classification. The process of the image classification algorithm is support vector machine (SVM) [14]. Support vector machine (SVM) is a supervised learning algorithm model that capable to analyze the data and identify patterns for the classification process. SVM algorithm is trying to find the most optimal hyperplane in the input space for the purpose of separating the two classes [15].



Support vectors are the inputs that intersect with the hyperplane. Where H1 is xi. w - b = +1 and H2 is xi. w + b = -1. The value of d + is the value of the shortest distance to the input points are positive and d- value is the value of the shortest distance to the point of negative input. Margin value is d + + d-. Marginal value should as much as possible to get the most optimal hyperplan.



4.4 Semi-global Matching

Semi-global matching is a method that successfully combines global and local concept of stereo to get the accuracy and suitability of pixel-wise with low runtime. This method has been used to image that has been fixed or not. In the end, epipolar lines will efficiently formed when the matching process [16]. In 2008, Hirschmller [8], developed the Global Semi maching method (SGM) and Mutual Information (MI) for stereo image data. Hirschmller implement DSM of the data in order to build aerial photographs of urban areas.

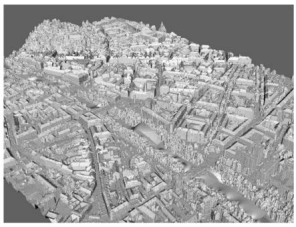


Fig. 8. DSM reconstructed

Figure 8 shows the results of the reconstruction of the Digital Surface Model (DSM). The limit of each object is very clear, but it depends on the amount of data. Over all, the qualities of the reconstruction results are excellent. Disadvantages of this method are requires high computational process. Accuracy of a dense stereo matching is important in 3D reconstruction. The hardest part is the object boundaries often not clearly visible and difficult to implement efficiently. It takes a method that can support pixel-wise matching to maintain the boundaries of objects clearly. But, this process should have low runtime, so it can support in real-time process.

5 The Proposed Method

Semi-global matching method in building the Digital Surface Model (DSM) shows that the method is able to achieve results with a high degree of accuracy.

Disadvantages of this method and the final requirement to obtain the proper buildings information can be solved by using the approach that we offer. The proposed approach diagram can be seen in fig 9 below. The proposed system consists of several activities, Analysis and design of its initial system is identification techniques for building extraction process, building height estimation and 3D reconstruction that will be used globally.

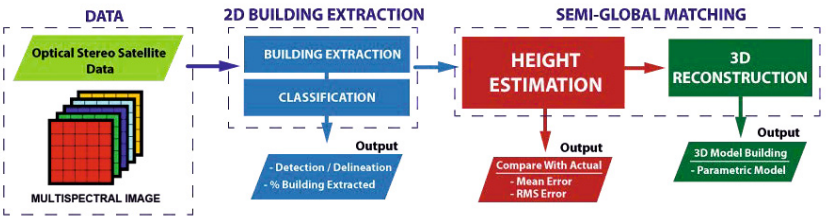


Fig. 9. The research process propose diagram

Implementation of building extraction system components. Where the system can recognize and separate objects contained vegetation and shadows in the image, so the objects contained in the image can be focused on building objects. The method at this stage is the NDVI method, SWI and image classification using SVM. Implementation the system components of building height estimation. The aim of the system is to build digital surface models (DSMs). The method implemented in this process is semi-global matching method. Implementation of system components for 3D reconstruction. At this stage, the system built a 3D model of buildings which exist in the satellite images. The results of 3D reconstruction of the building in the form of parametric models in which the end result has the necessary information such as the dimensions of the building and is able to be updated manually by the end user

6 Conclusion

This paper proposes a model for the 3D reconstruction of buildings automatically using satellite imagery data. In our approach, we propose to use a basic method Semi-Global Matching (SGM) with performed the detection and classification process. This approach aims to increase the accuracy of height estimation, faster computation time and low computational load. In further research, we will test the process of building extraction from 2D satellite imagery that begins with the process of matching data raw to reduce the data loss.

References

1. Ameri, B.: Feature based model verification (FBMV): a new concept for validation in building reconstruction. In: Proc. ISPRS Congress, pp. 24–35 (2000)
2. Cord, M., Jordan, M., Cocquerez, J.-P., Paparoditis, N.: Automatic extraction and modelling of urban buildings from high resolution aerial images. In: Proc. ISPRS Automatic Extraction of GIS Objects from Digital Imagery, Munich, Germany, pp. 187–192 (1999)
3. Vestri, C., Devernay, F.: Improving correlation-based DEMs by image warping and facade correlation. In: Proc. IEEE Computer Vision and Pattern Recognition, vol. 1, pp. 438–443 (2000)
4. Lin, C., Nevatia, R.: Building detection and description from a single intensity image. Comput. Vision Image Understand. 72(2), 101–121 (1998)
5. Noronha, S., Nevatia, R.: Detection and description of buildings from multiple aerial images. IEEE Trans. Pattern Anal. Mach. Intell. 23(5), 501–518 (2001)
6. Baillard, C., Zisserman, A.: Automatic reconstruction of piecewise planar models from multiple views. In: Proc. IEEE Computer Vision and Pattern Recognition, vol. 2, pp. 559–565 (1999)
7. Hashiba, H., Kamda, K., Tanaka, S., Sugimura, T.: Digital roof model (DRM)using high resolution satellite image and its application for 3D mapping of city region. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2003 (2003)
8. Hirschmüller, H.: Stereo Processing by Semi Global Matching and Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(2), 328–341
9. Izadi, M., Saeedi, P.: Height estimation for buildings with complex contours in monocular satellite/airborne images based on fuzzy reasoning and genetic algorithm. In: 10th Workshop Image Analysis for Multimedia Interactive Services, WIAMIS 2009 (2009)
10. Alobeid, A., Jacobsen, K., Heipke, C.: Building height estimation in urban areas from very high Resolution satellite stereo images. In: Proc. ISPRS, vol. 38 (2009)
11. Izadi, M., Saeedi, P.: Height estimation for buildings with complex contours in monocular satellite airborne images based on fuzzy reasoning. In: 17th IEEE International Conference Image Processing, ICIP (2010)
12. Unsalan, C., Boyer, K.L.: Multispectral Satellite Image Understanding: From Land Classification to Building and Road Detection. Springer, New York (2011)
13. Senaras, C., Ozay, M., Yarman Vural, F.T.: Building Detection With Decision Fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 6(3), 1295–1304 (2013)
14. Richards, J.A.: Remote Sensing Digital Image Analysis: An Introduction, 5th edn. Springer, New York (2013) ISBN: 978-3-642-30062-2
15. Berwick, R., Idiot, V.: An Idiot's Guide to Support Vector Machines. Advanced Computer Vision. University of Central Florida (2003)
16. Hirschmuller, H.: Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. In: IEEE Conference on Computer Vision and Pattern Recognition, USA (2005)

Intelligent Network

Spanning Tree Protocol Simulation Based on Software Defined Network Using Mininet Emulator

Indrarini Dyah Irawati and Mohammad Nuruzzamanirridha

Telkom Applied Science School, Telkom University, Bandung, Indonesia
indrarini@telkomuniversity.ac.id, mohammadnuruz@gmail.com

Abstract. Software Defined Networking (SDN) is a new networking paradigm [1] that separates the control and forwarding planes in a network. This migration of control, formerly tightly bound in individual network devices, into accessible computing devices enables the underlying infrastructure to be abstracted for applications and network services, which can treat the network as a logical or virtual entity. There are three methods for implementing SDN based network architecture, which uses Mininet emulator, Net-FPGA, and OpenFlow-based S/W switch [2]. In this research, simulation of Spanning Tree Protocol (STP) based SDN using mininet emulator. SDN is performed by OpenvSwitch (OVS) as a forwarding function, Ryu as OpenFlow Controller, and Mininet that installed on Raspberry-Pi. Then see the effect of the use of STP at Ryu Controller on network performance.The results of this research show that the simulation of SDN with OVS and Ryu controller can successfully runs STP function and STP on Ryu controller prevents broadcast storm.

Keywords: Spanning Tree Protocol, Software Defined Network, OpenFlow, OpenvSwitch, Ryu Controller, Raspberry-Pi.

1 Introduction

Nowadays, the development of computer network technology is growing rapidly. The growth will affect the increase of data traffic in the network. The traditional network architecture that integrates the forward plane and control plane into the same device cannot support these requirements. To solve this problem, the new network architecture method called Software Defined Network proposed. SDN is a new networking paradigm [1] that separates the control and forwarding planes in a network. The network can be dynamically managed depending on networking policies [3].

Mininet is a standard Linux computer network emulator that can create virtual topologies such as virtual hosts, switches, controllers, and link and support the OpenFlow protocol that can be used for computer networks based SDN simulation. Mininet is suitable for the purposes of research, development, learning, prototyping, testing, debugging, and some other functions just by using PC or laptop without any additional devices.

In this study, used mininet emulator to simulate a network topology SDN. SDN network consists of the forwarding plane in the form of OVS and control plane in form of Ryu Controller. Ryu Controller functions is as STP.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 395–403, 2015.
DOI: 10.1007/978-3-662-46742-8_36

2 Theory

2.1 OpenFlow

a. OpenFlow protocol provides standards for routing and delivery of packets on a switch. In the conventional network architecture, switch only works through packet forwarding to the appropriate port without being able to distinguish the type of protocol data transmitted. OpenFlow can perform the function of flow forwarding based network layer and manage centrally packet flow from layer 2 to layer 7 (flow granularity), so that the flow of packets in the network can be set up and configured independently. This can be done by making the algorithm and its forwarding rules in the controller which distributed to the switches on the network. OpenFlow architecture is shown in figure 1 that consists of OpenFlow Controller and OpenFlow Switch. OpenFlow Controller functions are controlled the path and managed OpenFlow switch. The example of OpenFlow controller are NOX, POX, Floodlight, Opendaylight and Ryu. OpenFlow switch as forwarding plane can implement by using hardware OpenFlow switch, Ethernet card that support FPGA, Mininet OpenFlow Switch Emulator and OpenFlow *S/W based* (example : OpenvSwitch that installed on Linux OS).

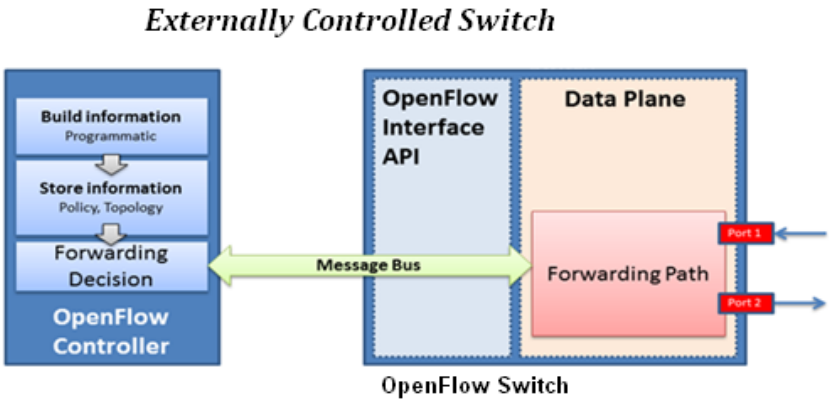


Fig. 1. OpenFlow architecture (Source: bradhedlund.com) [4]

2.2 Ryu Controller [5]

Ryu is a component-based software defined networking framework that shown in figure 2. Ryu provides software components with well defined Application Programming Interface (API) that make it easy for developers to create new network management and control applications. Ryu supports various protocols for managing network devices, such as OpenFlow, Netconf, OF-config, etc. About OpenFlow, Ryu supports fully 1.0, 1.2, 1.3, 1.4. All of the code is freely available under the Apache 2.0 license.

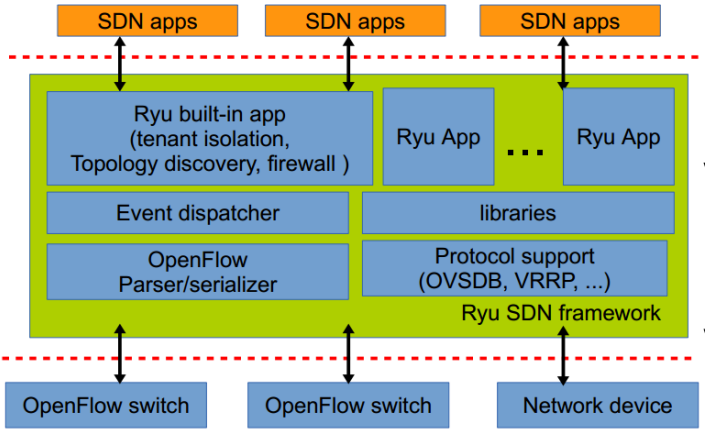


Fig. 2. Ryu SDN framework [6]

2.3 Spanning Tree Protocol (STP)

Spanning Tree Protocol is layer 2 protocol that defined by IEEE 802.1D standard. Spanning-Tree Protocol is a link management protocol that provides path redundancy while preventing undesirable loops in the network [7,8]. STP provides backup links between bridges and switches. The need for STP is based due to a broadcast storm. The broadcast storm caused a broadcast frame circling endlessly in the network due to the destination address in an unknown network. STP ensures only one path exists between anytwo stations.

3 Simulation Design

3.1 Network Design

In this research, using Raspberry Pi model B as a device that can be used as controller. Raspberry pi model B has specification with 512 MB of RAM, two USB port and 100 Mbps Ethernet port [9]. Design of network topology on this simulation is using mininet that installed on Raspberry Pi. The simulation is remote by secure socket shell (SSH) protocol on personal computer (PC).

The network simulation consists of 3 pieces of switchs are connected in mesh topology. Each switch is connected with a host. Network design is shown in figure 3.

3.2 SDN Controller

The SDN controller in this simulation uses Ryu controller software that installed in the raspberry pi with the Debian Wheezy version 3.12 Linux operating system. Ryu

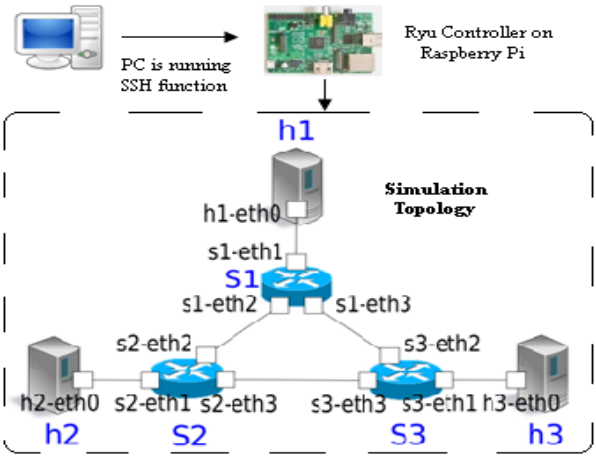


Fig. 3. Network topology

controller acts handle Spanning Tree Protocol (STP) on the network. Ryu SDN framework is programed with python programming language.

4 Result

There are two scenarios to evaluate the performance of Ryu controller. The first scenario is without using STP configuration on Ryu Controller and the second scenario is using the STP configuration.

To check the connection on the network is performed Packet Internet Grouper (PING) test from host 2 with IP Address 10.0.0.2/8 to host 1 with IP Address IP 10.0.0.2/8. PING operates by sending Internet Control Message Protocol (ICMP) echo request packets to the target host and waiting for an ICMP response. The ping message is used to reach the end-host.

4.1 SDN without STP Configuration

Based on the first scenario test, SDN without using the STP configuration, then the results are as follows:

- Broadcast storm occurs on the network that caused the ICMP packet is sent from the host with IP address 10.0.0.2 2 to host 1 with the IP address 10.0.0.1 is not sent (the status of host unreachable) as shown on figure 4. The figure is captured from host 2. The ICMP packet is cycling on the network.

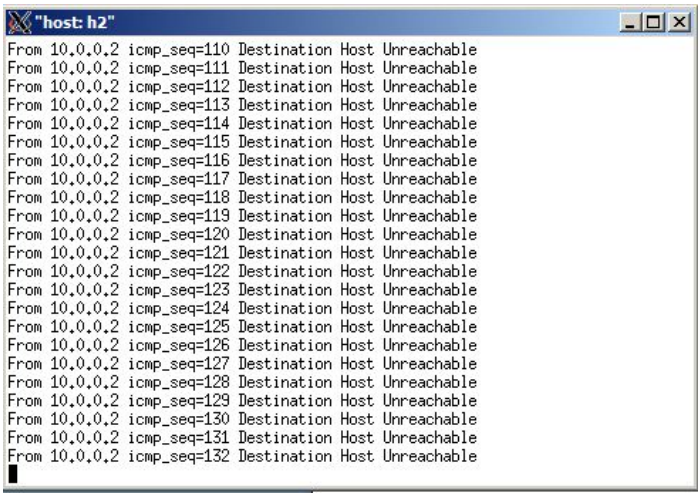


Fig. 4. Ping message from host 2 to host 1

- Results tcpdump on Switch 2 - eth1, there is only Address Resolution Protocol (ARP) Request seen repeatedly without ARP Reply as shown on figure 5. The send ARP function sends ARP request to obtain the physical address that corresponds to the specified destination IP Address (10.0.0.1).

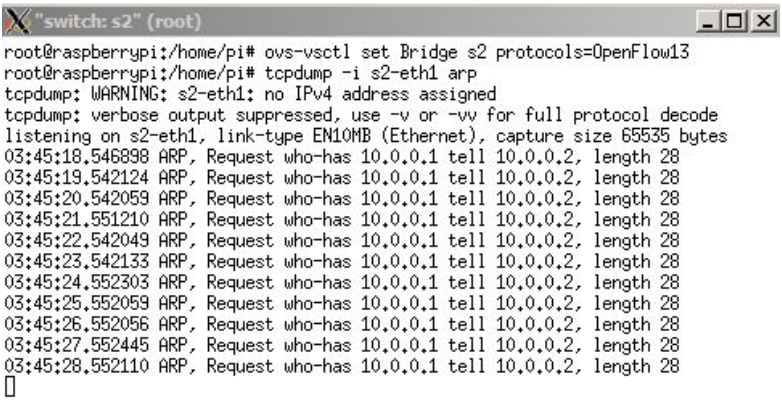


Fig. 5. *tcpdumpon* S2-eth1

- The network configuration is consisting of three switch (Switch 1, Switch2, Switch3), and three network segment (Switch 1, Switch2, Switch3). While Host 2 (IP address 10.0.0.2) transmits data to Host 1 (IP Address 10.0.0.1) then the traffic

- is broadcast from Switch 2 over to Segment 1; and it fails, then Switch 2 also broadcast traffic over segment 3. These switches would further create a flood. Hence, all switches keep sending and resending the traffic, eventually resulting in a flood loop or broadcast loop. Finally the network melts down, causing failure in all network links, which is referred to as a broadcast storm.
- Capture using wireshark as shown figure 6, only ARP Request seen repeatedly with information 'Who has 10.0.0.1 ? Tell 10.0.0.2, without an ARP Reply. That means the switch cannot find the destination address on the network.

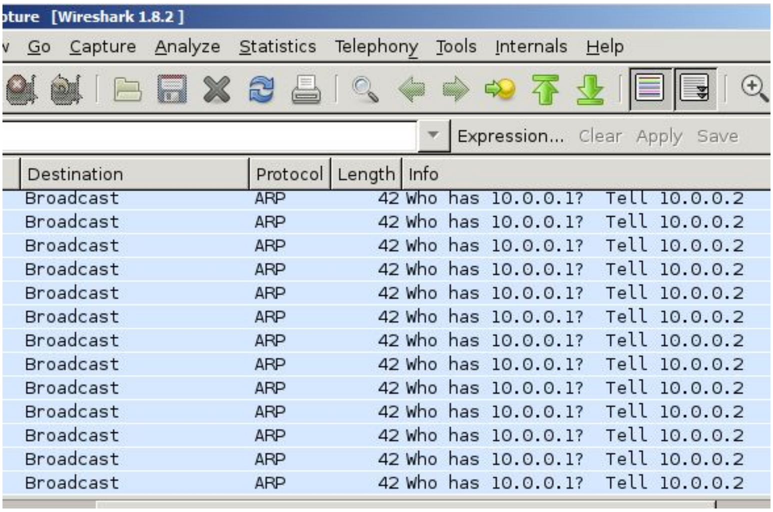


Fig. 6. Capture result from Wireshark

4.2 SDN with STP Configuration

- STP can be run by selecting the switch port state. When STP is enabling, the switch port starts from BLOCK state, and later changes to LISTEN state and LEARN state. BLOCK state is condition that the port does not participate in frame forwarding and discards frame received from the attached network segment. During the LISTEN state the port discards frames received from the attached network segment and it also discards frames switched from another port for forwarding. In the LEARN state , the port begin to process frame and starts update MAC Address table. FORWARD state is normal state that frame passed through the port. The port state can seen on the SDN controller as shown on figure 7.

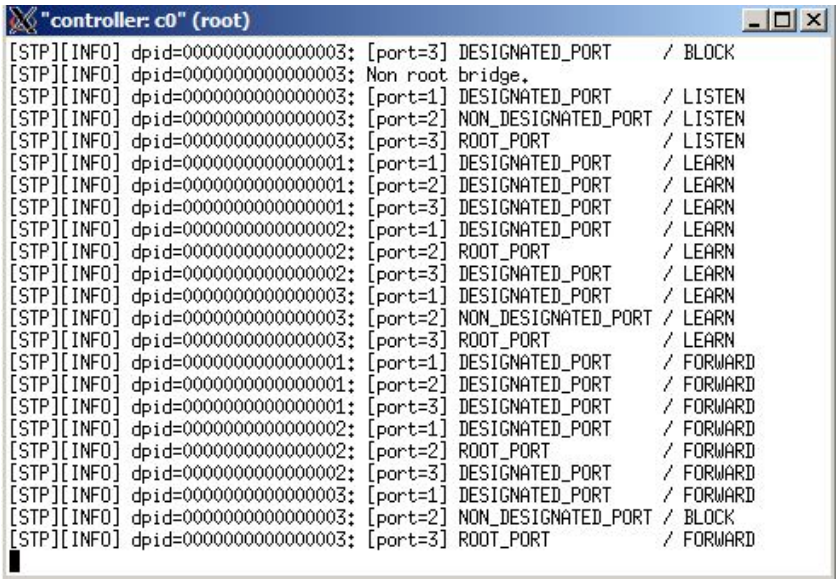


Fig. 7. Running STP on Ryu Controller

- The results of the ping message from host 2 IP address 10.0.0.2/8 to host 1 with the IP address 10.0.0.1/8 shown on figure 8. There is ICMP reply from host 2. The network still connected.

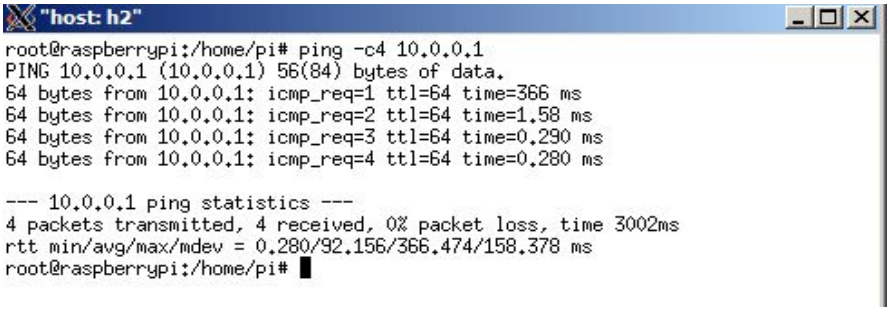


Fig. 8. Ping message from host 2 to host 1

- ARP Request and ARP Reply process is happened on Ryu Controller as shown on Figure 9. Ryu controller can handle the connection on the network and runs STP function. Capture from wireshark is shown on figure 10.

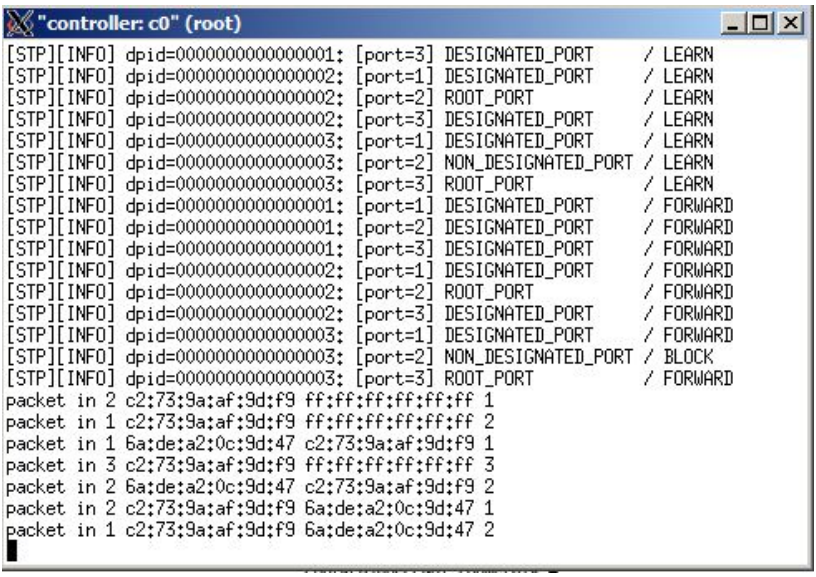


Fig. 9. ARP *Request*and ARP *Reply*on Ryu *Controller*

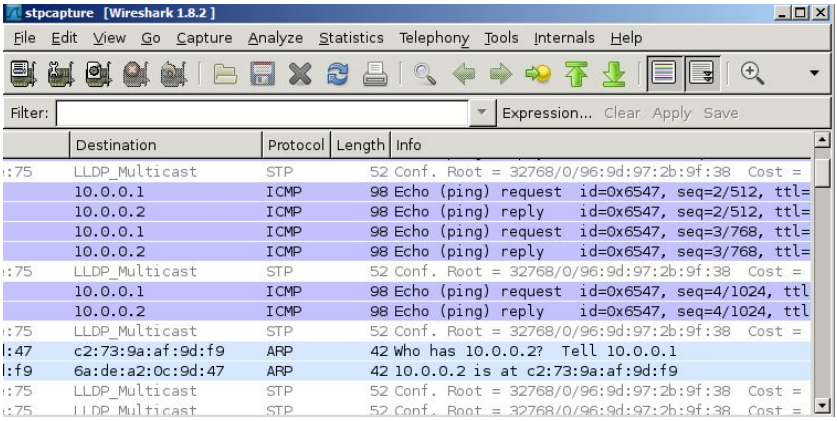


Fig. 10. Capture result from Wireshark

5 Conclusion

In this paper, the simulation of SDN with OpenFlow Switch and Ryu controller can successfully runs STP function. Ryu controller can perform STP function. STP stops flooding and prevents broadcast storm on the network. For future works, we will deploy the proposed SDN functionalities as STP controller using OpenFlow-based S/W switch to compare the result from mininet emulator.

References

What is OpenFlow?, http://archive.openflow.org/wp/learnmore/

Kim, H., Kim, J., Ko, Y.-B.: Developing a Cost-Effective OpenFlow Testbed for Small-Scale Software Defined Networking. In: ICACT 2014, Korea, pp. 758–761 (2014)

Sezer, S., Scott-Hayward, S., Fraser, P.K., Lake, D., Finnegan, J., Vijoen, N., Miller, M., Rao, N.: Are we ready for SDN? Implementation challenges for software-defined networks. IEEE Communications Magazine 51(7) (2013)

On data center scale, OpenFlow, and SDN, http://bradhedlund.com/2011/04/21/data-center-scale-openflow-sdn/

What is Ryu?, http://osrg.github.io/ryu/

Yamahata, I.: Ryu: SDN framework and Python experience. In: Pycon APAC 2013, Japan (September 2014)

802.1D MAC Bridges, http://www.ieee802.org/1/pages/802.1D-2003.html

Spanning Tree Protocol STP Tutorial, http://www.9tut.com/spanning-tree-protocol-stp-tutorial

Raspberry Pi Model B, http://www.rasberrypi.org/products/model-b/

Varnish Web Cache Application Evaluation

Justinus Andjarwirawan, Ibnu Gunawan, and Eko Bayu Kusumo

Informatics Department, Petra Christian University, Surabaya, Indonesia
{justin,ibnu}@petra.ac.id,
m26410006@alumni.petra.ac.id

Abstract. Websites today have static and dynamic contents that concern many people about the performance. People will leave or not returning the website because of loading time that is taking too long. One way to speed up a website access is by utilising web applications using Varnish cache. Varnish will store static content from a website in the memory. Static data is data that changes infrequently or rarely updated. A front end for web administrators was built for Varnish application so that the configuration can be done easily. The configuration that will be simplified through the front end of this Varnish application is detecting SQL injection and Cross Site Scripting (XSS), block files and folders, HTTP header manipulation, detects the original file whether it is deleted or blocked and error handling configuration. The front end application is implemented on Linux platform. Varnish application was tested using ApacheBench where the number of requests from the client and the response time become the main parameters of the test.

Keywords: web cache, reverse proxy, varnish.

1 Introduction

Fast access to websites is everyone's demand at this time. Website visitors leave the website when loading is too long. Websites often have additional content that do not support the main information of the website. This additional content will slow down the access to the website. On a website there are static and dynamic content. Static content is content that is updated infrequently or rarely changes. Dynamic content is content that is frequently updated or there is always a change. Static content such as header, sidebar and footer while dynamic content such as news content, gallery or video. The static content can be stored in a memory so that dynamic content data access to the server faster. The data stored in this memory will speed up static content access on the website so that if a user visits a lot and load dynamic content, server access data will be lighter.

Another thing that becomes a problem is if the main server where the data stored is far from the location of a visitor. The time required to transfer the necessary data will be longer. The data in this memory is needed to accelerate the transfer of this data. For example, someone has a web server in the U.S., but the majority of website visitors are from Indonesia. Data access will be a little longer because the web server

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 404–410, 2015.
DOI: 10.1007/978-3-662-46742-8_37

is away from the users. Data in the memory also helps keep some frequently accessed data.

Web caching is a solution to this problem by using a web cache called Varnish. With Varnish, one can store all the data information from the primary server to the Varnish. Static data such as images, videos and other things related to the website may be stored in this Varnish. Web admin can also set the Time To Live (TTL) of each website fragment by using the concept of Edge Side Includes (ESI). Website content that is rarely updated like the header or footer can be set up its time to live for example like one month. During one month, Varnish automatically update its cache site. Content such as news or advertising, can be set e.g. for 1 hour to be updated. Distribution of the update time is beneficial to reduce data access to major web servers and data received the most recent data.

If a website is accessible to many visitors the loading will be quite long. A Varnish cache site is to help increase the speed of the website access. It is expected to help the problem on the server load and access data through the use of Varnish web cache. Implementing Varnish cache in this research is to define how significant it is to use on highly loaded web servers.

2 Theoretical Background

Varnish cache is a web accelerator application, also known as Hyper Text Transfer Protocol (HTTP) reverse proxy. Varnish was first introduced in 2006 by Poul-Henning Kamp. Reverse proxy is a proxy that is in the front-end of web servers to act as a cache. Varnish work on the front end of an HTTP server and can be configured to cache every website content. Varnish's main working principle is to store the data of a web page in memory, thus reducing the load on the web server loading the same page (Winkler 2012). Here the web application without caching process can be seen in Figure 1:

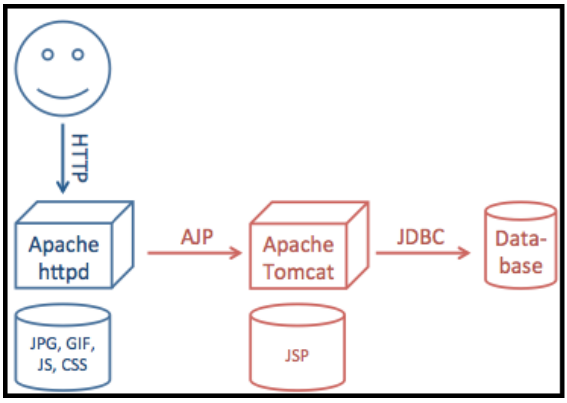


Fig. 1. Web application without cache

The three main components of the web application is Apache HTTPD, Apache Tomcat and database. Apache set static requests for images, scripts and others. In addition there is a forward requests for pages Hyper Text Markup Language (HTML) to application servers such as Apache Tomcat and then request to the database. After successfully getting the information request, it will be returned to Apache and then to the user. The red component is the slowest component in its performance due to the long process of searching data for this component. (Winkler 2012)

If researchers apply the varnish on a website, Varnish HTTP accelerator will serve as the store (cache) a copy of each page of the website. When a user accesses the website back, Varnish will provide a copy of the data of the Apache server. Varnish will greatly assist access to a website with a system of storing data in the cache memory of this site. The following implementation of Varnish web cache can be seen in Figure 2:

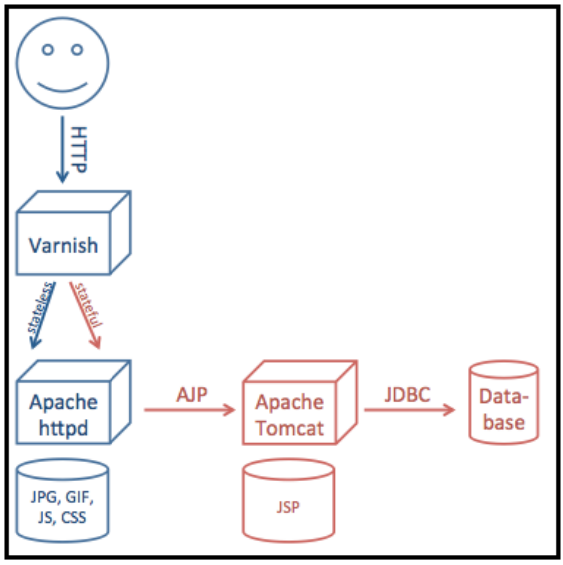


Fig. 2. Web server with Varnish web cache

Varnish is a web cache that will be divided into two: stateless and stateful. Stateless is data with static content, while stateful is data with dynamic content. Varnish only serves requests stateless whereas for stateful request will be returned to Apache web server.

VCL (Varnish Configuration Language) is a programming language used in the Varnish web cache. VCL syntax is similar to programming in C language.

3 Research Model and Hypotheses

3.1 System Analysis

Varnish that has been installed can be accessed via the default port 80. Varnish will be at the front end admin so that requests data going through Varnish will be checked first before heading to Apache. Request a stateless data will be managed by Varnish while stateful data request will be returned to Apache. If there is a request then the request will be checked by Varnish whether it is already cached. If not cached, then requests will be forwarded to the Apache. If there is a page-request it will weigh on the performance of Apache. Varnish will assist in arranging Apache request the same page with the system cache in-memory store. The system after using Varnish can be seen in Figure 3:

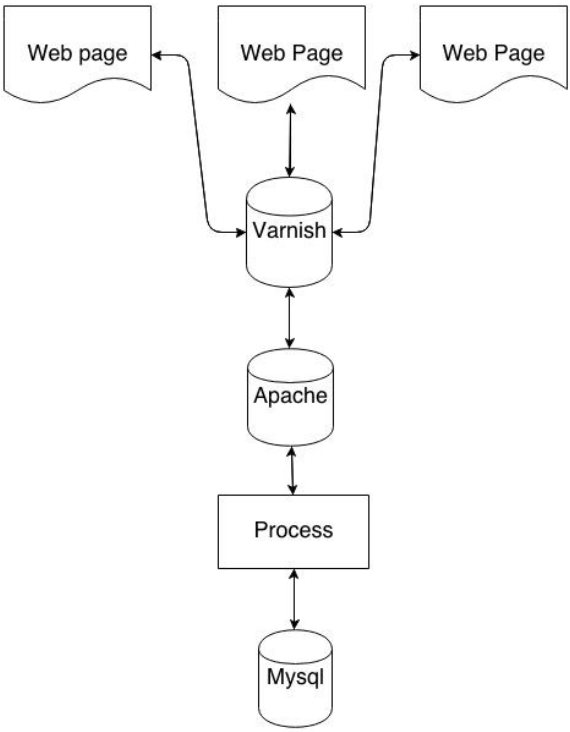


Fig. 3. System after using Varnish

The analysis test bed is to separate static and dynamic content to decide which part of the web to be cached by Varnish. The separation needs to be done manually by the web administrator.

3.2 Design of Experiment

Three websites were developed for the test. One site has 10 kilobytes of static content, the second site with 4 kilobytes and third site 15 kilobytes with mixed static and dynamic content with PHP backend. These three websites were placed locally on the same machine as the benchmark tool ApacheBench.

4 Web Cache Performance Test

Gnuplot is used for the visualization in the form of graphs. In Figure 4 it can be seen that the response time benchmarks without Varnish starts to weigh when there are 40 incoming requests. While the response time benchmarks via Varnish remains stable during the 40 incoming request. Response time is the longest of the benchmark, without Varnish is 5048ms and reponse time longest of benchmarks with Varnish is 264ms with a maximum of 100 requests. Graph on figures are generated using Gnuplot [2].

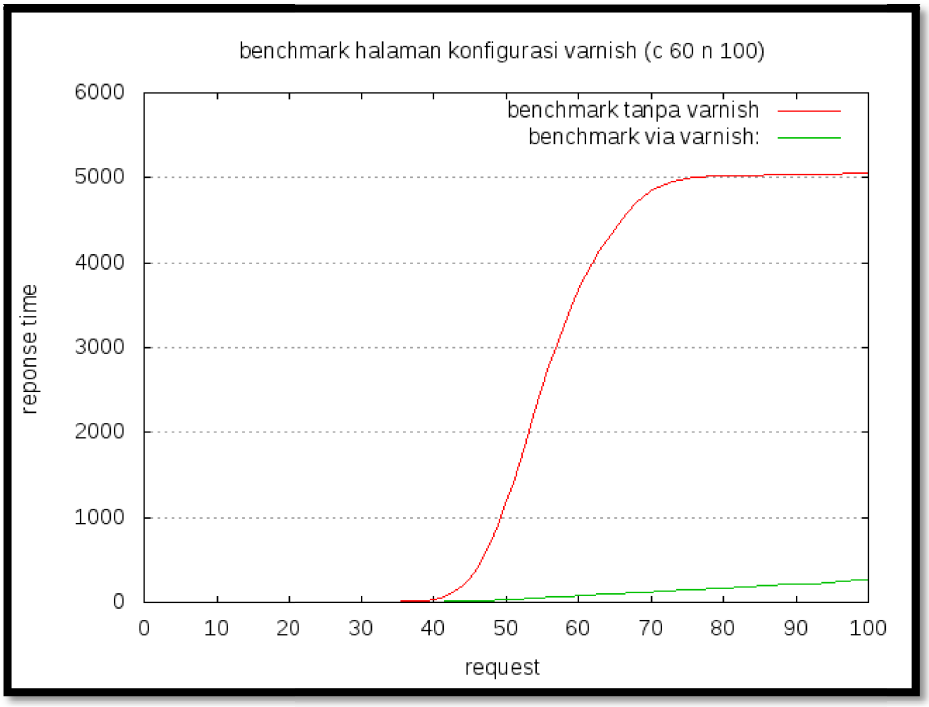


Fig. 4. Benchmark test

Using ApacheBench for benchmarking, setting a concurrent users value to 296 and number of requests 300, the following table shows the connection time (ctime), processing time (dtime), total time (ttime) and waiting time:

Table 1. Varnish benchmark test

starttime	seconds	ctime	dtime	ttime	wait	id user
Wed May 28 10:42:45 2014	1401248565	0	1	1	1	1
Wed May 28 10:42:45 2014	1401248565	0	2	2	2	2
Wed May 28 10:42:45 2014	1401248565	0	2	2	2	3
Wed May 28 10:42:45 2014	1401248565	0	3	3	3	4
Wed May 28 10:42:45 2014	1401248565	8	2	11	1	5
Wed May 28 10:42:45 2014	1401248565	8	2	11	2	6
Wed May 28 10:42:45 2014	1401248565	8	2	11	2	7
Wed May 28 10:42:45 2014	1401248565	8	2	11	2	8
Wed May 28 10:42:45 2014	1401248565	8	3	11	2	9
Wed May 28 10:42:45 2014	1401248565	8	3	11	3	10
Wed May 28 10:42:45 2014	1401248565	9	2	11	2	11
Wed May 28 10:42:45 2014	1401248565	8	3	11	3	12

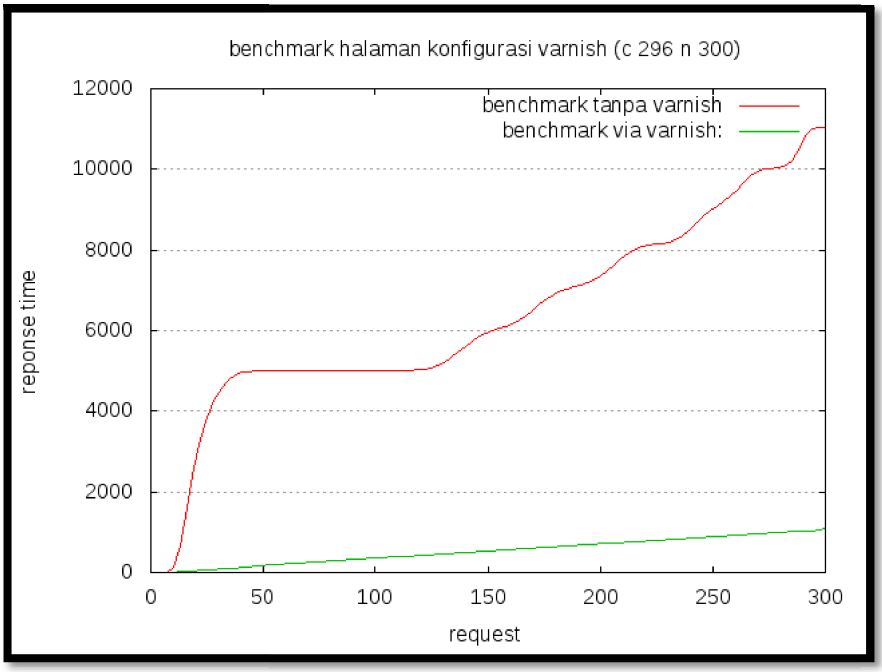


Fig. 5. Benchmarking Varnish

The table above is shown as figure in Figure 5:

The longest response time on the 300 users configuration without using Varnish is 11047ms compared to 1087ms with using Varnish.

The remaining tests with various numbers of concurrent users and requests showed similar conditions from Figure 5.

Varnish has a built-in script feature called VCL (Varnish Configuration Language) for making its configuration and also a security feature to prevent XSS (Cross Site Scripting).

5 Conclusion

The evaluation showed that Varnish is accepted significantly on static contents of websites. Not very useful for websites which have frequent updates or very dynamic content changes. Varnish still need cache clearing mechanism to delete cached contents in order to retrieve updated content after a period of time. It can also be concluded that Varnish is very recommended to implement on large scale web servers with high concurrency load.

References

1. Christian, W.: Ultra-Performant Dynamic Websites with Varnish (2012), http://blog.mgm-tp.com/2012/01/varnish-web-cache/ (accessed November 18, 2013)
2. Gnuplot, Gnuplot Homepage (2013), http://gnuplot.info/ (accessed November 18, 2013)
3. Silicon-Press. Web Caching: What is Web Caching?) (2013), http://www.silicon-press.com/briefs/brief.webcaching/brief.pdf (accessed November 18, 2013)
4. Varnish Varnish Book (2012), https://www.varnish-software.com/static/book/ (accessed November 18, 2013)

DACK-XOR: An Opportunistic Network Coding Scheme to Address Intra-flow Contention over Ad Hoc Networks

Radha Ranganathan[1], Kathiravan Kannan[2], P. Aarthi[3], and S. LakshmiPriya[4]

[1] Asst Professor, IT Dept, Easwari Engineering College, Chennai, India 600089
[2] Dean, Easwari Engineering College, Chennai, 600089 India
[3] Software Engineer, Philips India Ltd, Bangalore, 560045 India
[4] Student, Arizona State University. Tempe, AZ 85287-1003
{radhupriya@yahoo.co.in,
{kathirraji,aarthi.npk,slplakshmipriya}@gmail.com}

Abstract. Network coding is a novel technology that exploits the intrinsic broadcast nature of wireless media, to significantly reduce the number of transmissions (hops). Our primary focus is to minimize intra-flow contention by exor-ing TCP-DATA and TCP-ACK packets belonging to the same TCP flow, ensuring that the packets are never delayed in the process. Network coding always comes with the overhead of intermediate nodes having to buffer packets so as to successfully perform decoding. This requires the nodes to maintain large buffers. We also propose a new technique in which we retain only the last delivered packet in the buffer. However, at times when the same node gets access to the medium repeatedly, keeping only the last sent packet in the buffer will not suffice. Hence we introduce a rate based Cross-layer Transport Solution (CLTSP) that inserts a delay interval (called out -of-interference delay) between two consecutive packets transmitted. This reduces intra-flow contentions by leaps and bounds. Our unique combination of using a Network Coding technique along with a suitable TCP variant improves the throughput gains with significant improvement in handling medium contention by reducing the number of transmissions.

Keywords: Network Coding, Intra-flow Contention, DATA-ACK in TCP, Ad Hoc Networks, EXOR coding, Pseudo-Broadcast.

1 Introduction

With gadgets galore, people are living a highly networked life these days. The Internet and the cell-phone network has literally transformed the life of every single individual on this planet. The growing number of users coupled with the increasing demand for newer technology is the principle factor that drives the network scientists to think beyond the horizon.

In the Internet, the responsibility for directing data traffic lies with special-purpose devices called routers. Internet service providers monitor the flow of traffic across their networks and, if they spot congestion, revise the routers' instructions accordingly. With

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 411–423, 2015.
DOI: 10.1007/978-3-662-46742-8_38

the cell network, two people a block apart could be having a phone conversation, but they aren't directly exchanging data. Rather, they're sending data to a cell tower that determines what to do with it. In recent years, many network scientists have turned their attention away from centralized networks — such as the Internet and the cell-phone network — and toward ***ad hoc networks***, wireless networks formed on the fly by.

The Transmission Control Protocol is a well-known de-facto protocol in developing today's internet. Due to its wide acceptance and deeper understanding, it is desirable to extend and adopt its functionalities in Wireless Networks also. But studies have shown that TCP performs poorly in MANETs [1]. The reasons can be attributed to the typical behavior of the nodes in the wireless medium - mobility, high bit error rate, unpredictability, contentions, long connection times, etc. Some TCP-Specific problems identified over Mobile Ad-hoc networks[2]:

- TCP misinterprets route failures as congestion
- TCP misinterprets wireless errors as congestion
- Intra-flow and inter-flow contention reduce throughput and fairness
- Delay spike causes TCP to invoke unnecessary retransmissions
- Inefficiency due to the loss of retransmitted packet.

We propose to focus our work towards reducing the intra-flow contention that has significant impact on the throughput of the network. Of the various means of alleviating this kind of medium contention pertaining to a single connection, our work uses a new kind of Network Coding technique to address the issue.

The core idea of Network Coding is to mix several packets together for transmission, thereby greatly increasing the amount of packet information transmitted in a single hop. The intermediate routing nodes no longer just store and forward packets. They also process the packets before transmission. This processing done at the relay nodes helps to reduce the number of packet transmissions at the MAC layer and consequently the number of times a node contends for the medium.

2 Related Work

Network Coding techniques are being touted as networking's next revolution [3] and it is considered as a paradigm shift in data transportation across networks. No wonder that the academia and the Industrial giants are looking forward to embrace this technology in the existing network infrastructure systems. Since the time when Network Coding Theory [4] stemmed out of the lineage of Information Coding theory, lot of pioneering work has been carried out to exploit the Network Coding paradigm with multi-faceted approaches. A variety of coding techniques were adopted for different purposes. Network Coding has huge potential impact in multicast applications as illustrated in the works of [5], [6]. The Unicast traffic scenario has been considered in the [7], that exploits physical layer capabilities and in COPE [8] wherein Medard et al. have proved the prowess of the Network Coding abilities with a practical test-bed implementation. This work looks out for effective coding opportunities to forward multiple packets in a single transmission. However, this demonstration could not produce desired levels of throughput gains in TCP end-to-end connections. Sundararajan et al. [9] proposed a intra-flow random network coding

scheme with a new interpretation of ACK packets with the concept of Degrees of Freedom. But this work failed to bring out the fullest potential of the wireless network coding as it employed coding operations only in the end hosts and not in every relay node. Chuan Qin et al. [10] proposed a first-of-a kind system to integrate both inter and intra flow network coding operations in the wireless medium. But the suitability of this project in TCP connections is quite ambiguous. A Mac-level Network coding scheme named – Piggycode [11] had exclusive focus on the intra-flow packet coding mechanism which significantly improves TCP performance. Mario Gerla et al. used this work in their ComboCoding technique [12] with an additional timer mechanism to the Piggy code system. Both these works emphasize the need for an almost infinite buffer for effective operations Our proposed work, DACK-XOR, which is DATA-ACK EXOR Network Coding Scheme‖ is to enhance the pure coding-decoding procedures, i.e. XOR- network coding in a TCP connection to mix TCP DATA and TCP ACK packets and strive to achieve better throughputs and reduce the stress on the medium contention issue by restructuring buffer systems and Transport Layer mechanisms.

3 Network Coding - A Stress on the Buffer?

Network Coding mandates the maintenance of an infinite decoding buffer to store packets that have already been sent, in order to perform successful decoding. This comes as a huge overhead in terms of the buffer space required. Also, this introduces a significant delay in the decoding process because the packet needs to be searched for in the infinite buffer. Therefore, we resort to a technique which retains only the last packet delivered, in the buffer. This takes the stress off the buffer to a large extent.

However the above mentioned technique works only if there is a strict fairness in medium access. So we make use of a rate based TCP, to ensure that the same node does not repeatedly get access to the medium. This TCP variant, in addition to reducing intra-flow contention also sees to it that packets do not get clogged at a specific node. Fig. 1 shows a sample schematic view of our DACK-XOR ing Scheme of Network Coding.

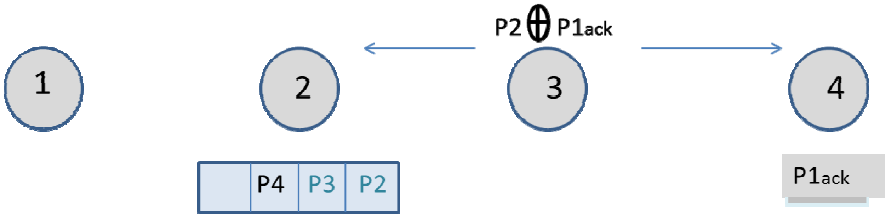


Fig. 1. Network coding illustration with new proposal

4 System Architecture

The network coding logic is integrated into the MAC layer. Packets coming from the upper layers are lined up in the outgoing queue, waiting to be transmitted. Packets that are received from outside are also queued up, waiting to be forwarded. The network coding layer looks for a coding opportunity by checking if a TCP-DATA residing in

the queue can be exored with a TCP-ACK and vice-versa. If the look up is successful, the two packets to be exored are sent into the encoder where the TCP-ACK packet is appended with an appropriate number of zeroes so as to be exored with TCP-DATA packet. The coded packet is then broadcast. A copy of the sent packet is kept in the decoding buffer (packet pool). When a coded packet is received, it is sent into the decoding unit. Using the unique packet identifier as the search key, the native packet is retrieved from the packet pool. This known packet is exored with the received packet to obtain the original packet. Fig.2 is our Generic node architecture that supports DACK-XOR Network Coding technique.

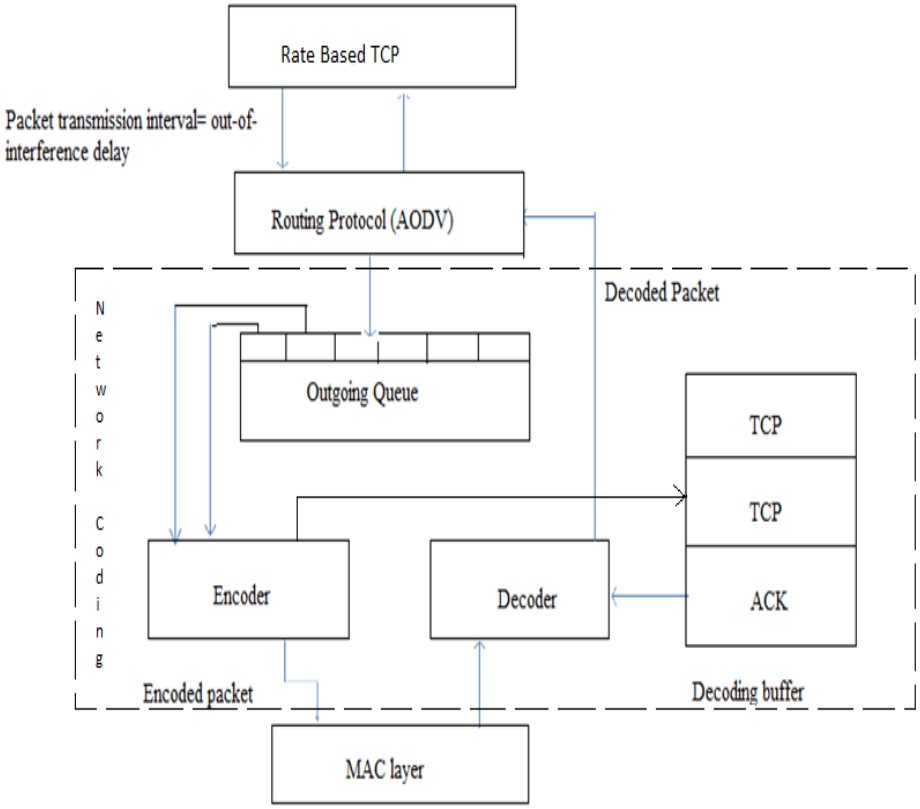


Fig. 2. Generic Node Architecutr to support DACK-XOR

5 Module Description and Design

Our implementation aimed at lessening the medium contention within a specific TCP flow by reducing the number of transmissions of packets across the medium involves three different modules. The first one is the heart of the Coding-Decoding part of the Network Coding Module. The second one is focused on improvising the Encoded packets' transmission by using a pseudo-broadcast mechanism instead of the usual

Broadcasting approach used in many of the Network Coding works, as this offers better reliability and feedback mechanism. The third and the last module works on refreshing the packet pool appropriately to reduce the stress on the buffer. This work looks out for effective coding opportunities to forward multiple packets in a single transmission. However, this demonstration could not produce desired levels of throughput gains in TCP end-to-end connections.

5.1 Piggy Code Network Coding

In this module, we implement network coding which piggy-backs data and acknowledgement in a single packet transmission to two different receivers in a single TCP connection, assuming that TCP DATA and ACK packets travel in the same path but in opposite directions. Here, we make use of decoding buffers of infinite size to store a copy of all the sent packets.

1) Opportunistic Encoding Process:
When we de-queue the packet at the head of the interface queue at the relay node, we check if there are coding opportunities, by searching for a packet of opposite type. We are said to arrive at a coding opportunity when a packet of opposite type with interchanged source-destination pair addresses and belonging to the same flow are identified in the queue. The Algorithm for Interface Queue Management is given in Table. 1.

Table 1. Algorithm for Interface Queue Management

```
INPUT: Q, the Packet Queue at Link Layer
OUTPUT: Packets to be sent to Down Target – MAC
PROCEDURE:
Initialize chained:=0,
     P:=head of Interface Packet Queue ,
CurrPtype:= P • type, CurrConnID:=P • flowID,
CurrSrc:=p • IPsrcAddr, CurrDst:= IpdstAddr
OppType:= (CurrPtype==TCP-DATA)? TCP-ACK : TCP-DATA
If CurrPtype == TCP-DATA or TCP-ACK and chained ==0 then
  pkt:= P
  foreach pkt:=pkt • next
    if pkt!=NULL then
      if pkt•type==OppType and pkt • flowID == CurrConnID
      then
        if pkt•IPsrcAddr==CurrDst and pkt•IPsrcAddr ==
        CurrSrc then
          chained:=1
          remove(pkt)
          p • next:=pkt
          break
        endif   endif endfor endif
```

If the queue lookup is successful, the two packets to be exored are chained together. The chained packets are de-queued together from the Link Layer Packet Queue and transmitted to the down-target, i.e., the MAC layer.

It is in the MAC layer wherein we perform the actual Network Coding Operation. We EXOR the payload fields of the MAC layer and put them up in the new EXOR-ed packet. The packets arriving at the MAC layer to be passed on to the Wireless Physical Layer are stored in the Packet Pool for decoding purposes. They pool is refreshed at appropriately, as explained in this section later.

We then include a CODED field in the Network Coding header (NCHeader) of the new packet to indicate that the packet is coded. We also insert the sequence ids of the two packets that have been coded. The new packet is now sent further down for transmission. The algorithm used for the Encoding of suitable packets opportunistically, is illustrated in a flow-chart in Fig. 3.

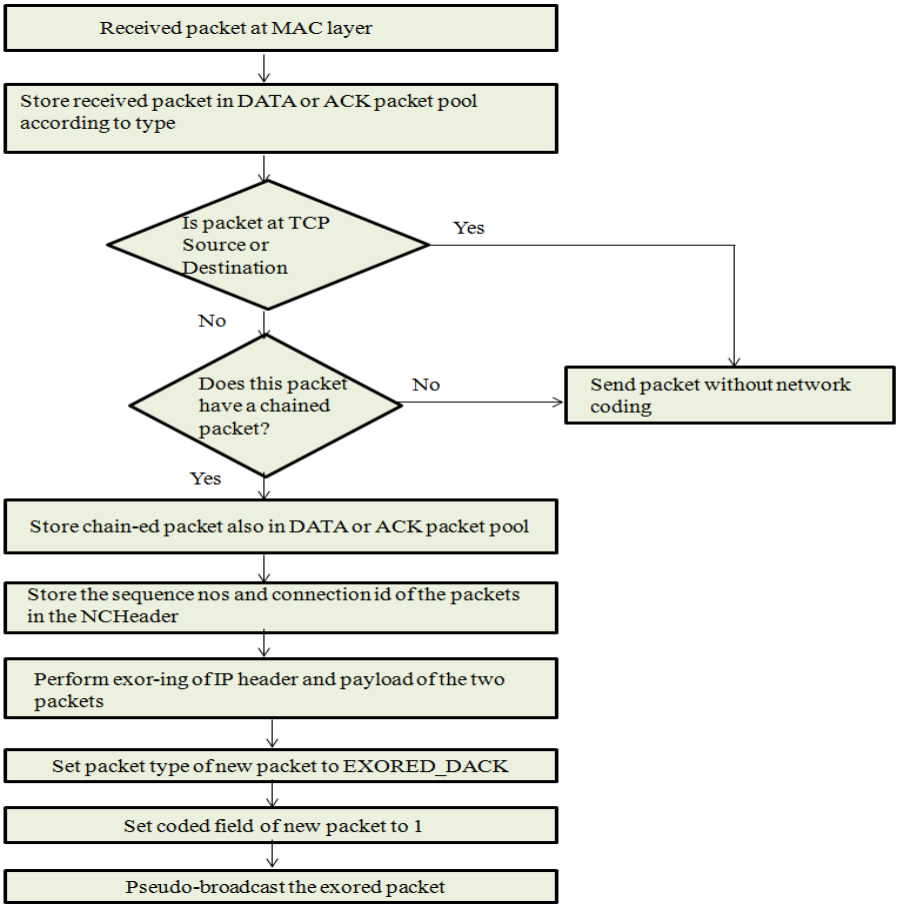


Fig. 3. Network Coding - Encoding Process flow chart

2) Decoding Process:
When a coded packet is received, we check if one of the packets inside the coded packet has already buffered been in the pool. This lookup is performed using the packet identifiers as the search key. If the un-coded packet is found, it is exored with the received packet in order to retrieve the original packet. Fig. 4 depicts this process as a flow-chart.

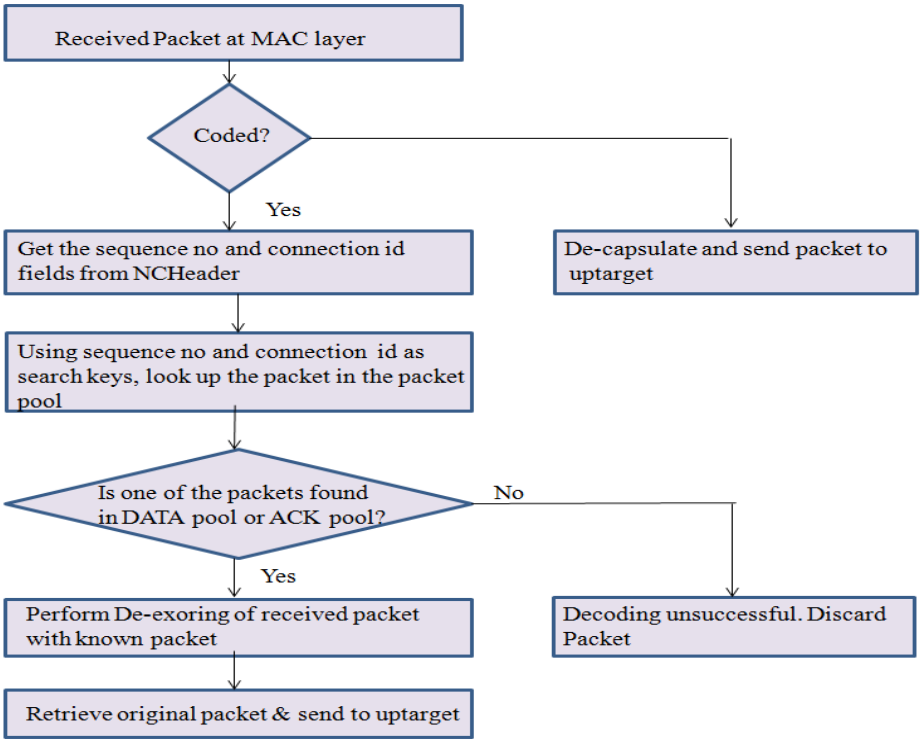


Fig. 4. Decoding Process flow chart

5.2 Pseudo – Broadcast of Exored Packet

The broadcast of exored packets might lead to collisions in the medium due to lack of RTS-CTS exchange in 802.11. This results in the TCP source having to retransmit the packet. Hence we resort to pseudo-broadcast of the coded packet. The coded packet is now unicast to only one of the two intended recepients by including the receiver address in the MAC header. The other receiver's address is stored in our new NCHeader. Since any node within the interference range of the sender can receive packets that are not destined for it, the node receives the packet and checks the NCHeader to see if it is the other intended recepient. If yes,it performs de-exoring of the packet. Pseudo-broadcast decreases the collision rate to a considerable extent

5.3 Coding Buffer Enhancement – Last Delivered Packet

In this module, we restructure the Piggy Code Buffer of every node to retain only the last packet delivered. This greatly reduces the buffer look-up time for decoding packets at the receiver side. In order to be able to perform decoding successfully, this needs to be done in conjunction with the third module "Rate based transport solution".

5.4 Implementing Rate-Based Transport Solution

We make use of a rate-based Cross-Layer Transport Protocol (CL-TSP) that enforces strict fairness in medium contention, by making sure that the same node does not get repeated access to the medium. This is achieved by exploiting the spatial reuse of the wireless channel. The CL-TSP inserts a delay in between successive packet transmissions. The delay is calculated by averaging the out-of –interference delay in the forward and reverse path. Essentially, once the CL-TSP layer of a node transmits a packet, it has to wait for a specific amount of time before it can transmit the next one. This prevents intra-flow contentions among successive packets.

We deliver only one data packet at a time from transport layer of source to prevent intra flow contention among them. The time interval between the successive deliveries should be calculated such that it should not be large which may otherwise underutilize the bandwidth of the network and it should not also be small which can otherwise lead to contention among them. If packet 1 is delivered at time t1 and packet 2 is to be delivered at time t2, then the time interval (t2-t1) must be carefully selected by considering the above factors.

We can say that the time interval between the deliveries of successive packets at the transport layer can be the four hop transmission delay which is named as out of interference delay. So, (t2-t1) should be at least the out of interference delay to prevent contention between pkt1 and pkt2.

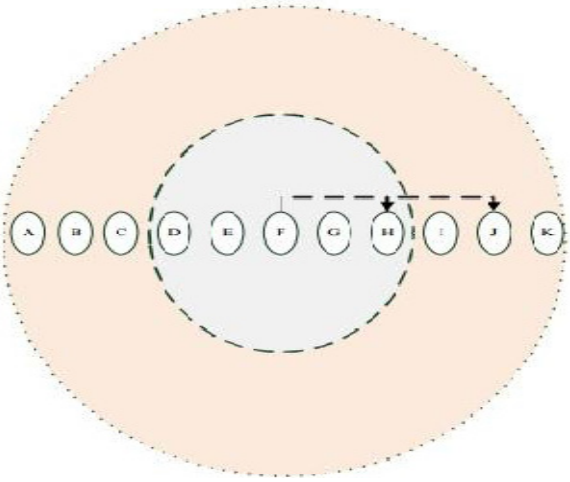


Fig. 5. Analysis of number of hops on out of interference range

We may also argue that the out of interference delay cannot be always represented as four hop transmission delay. It will increase when the nodes are closely placed. For example, we consider Fig.6. Where in the distance between the nodes are taken as 100m (closely located). The transmission range of node F covers node G, H and its interference range covers up to node K. The routing protocol like AODV chooses shortest path which is based on hop count to reach the destination. The next hop for node F is chosen as node H not the node G. The path may be chosen as node F ->node H->node J ....So, we assume that four hop transmission delays can represent the out of interference delay in most cases.

We need to analyze whether this four hop out of interference delay will be sufficient to determine the inter packet delivery period at CL-TSP source. The transport layer should be reliable in nature, the CL-TSP source should receive acknowledgment for every data packet from the end receiver. Also, the forward and reverse path of the transport connection is same in most of the network. In this scenario, if four hop transmission delay alone is considered in calculating the inter packet delivery, it will lead to severe contention with the reverse acknowledgments. So, we need to consider the four hop transmission delay of both data and acknowledgment packets in determining the inter packet delivery period. The total delay for a packet i can be measured as given in Eq. 1

Total delay(i) = max(interference delay values recorded)

We will be taking into account the total delay of recent n packets for calculating the exponential mean which will be represented as the inter delivery time period between the next two packets.

6 Experimental Results

We evaluate the performance of our DACK-XOR Network Coding in a phased manner to emphasize the significant performance improvements gained through our work. The specialty of our work lies in the fact that the modularized parts can be incrementally deployed in real time scenario. To prove this point, we have chosen to evaluate our work with four different variants:

- IEEE 802.11 at MAC layer + TCP New Reno at Transport layer protocol and this combination is named as " original MAC" in the comparison charts.
- IEEE 802.11 at MAC layer + PiggyCode Network Coding Module +TCP New Reno at Transport layer protocol and named as "Piggy Broadcast" in the comparison charts.
- IEEE 802.11 at MAC layer + Piggy Code Network Coding Module with Pseudo broadcast + TCP New Reno at Transport layer protocol and named as "Pseudo Piggy" in the comparison charts.
- DACK-XOR integrated module at MAC + rate based CLTSP at transport layer and named as "Integrated TSP" in the comparison charts.

The finally integrated DACK-XOR package comprises all the modules together – PiggyCode Network Coding with Pseudo Broadcast implementation and a Rate-based

CLTSP solution - to achieve the ultimate objective of alleviating Intra-flow Contention. We have used Network Simulator 2.34 for our Experimental Analysis. Our chosen Network Scenario is a Linear-chain topology with 8 nodes (7 hops). The distance between any two contiguous nodes is set to 220 meters and the transmission range of each node is 250 meters. We consider a single TCP connection with Node 0 as the source and Node 7 as the sink. The various Simulation parameters for our Network Scenario are listed down in Table 2 as shown below.

Table 2. Simulation Setup

Application Type	FTP
Data Rate	Discrete variants from 1Mbps to 5 Mbps
Basic Rate of 802.11 MAC	1Mbps
IFQ Capacity	50
Queue Type	DropTail / PriQueue
Propagation Model	TwoRayGround
Antenna Type	Omni Antenna

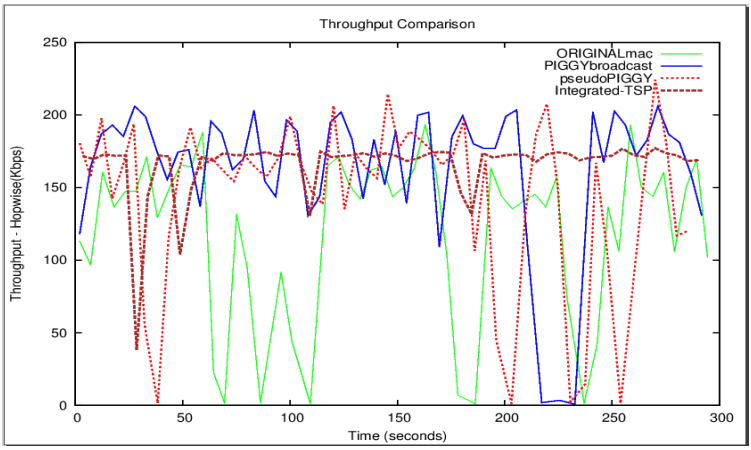


Fig. 6. Throughput variance across time for our Network under different Protocols

6.1 Throughput vs Time

We have shown the performance of our network-coding variants by measuring the throughput at regular intervals of time during the entire simulation. It can be seen from Fig.6 that throughput of our integrated network coding solution is consistently high as opposed to the original scheme without network coding in which the throughput drops to zero many a times.

6.2 Number of Hops vs Throughput

Having shown the continuous throughput variation with time, we next evaluate the same parameter with respect to the number of hops in the network scenario. The graph in Fig. 7 shows the average throughput obtained for varying number of hops. The effectiveness of our scheme can be realized only for a number of hops greater than 5 when the effect of network coding becomes pronounced. Also, the out-of-interference delay comes into play only when the number of hops is greater than 4.

6.3 Throughput under Different Data Rates

Here we measure the average throughput of the four schemes at different data rates such as 1, 2, 3, and 5 Mbps. It can be seen that the effectiveness of network coding cannot be felt at 1Mbps data rate. This is because in spite of TCP trying to inject as much packets as possible, the available bandwidth is less and the number of packets travelling across the medium is also less as explained in [5]. So, only a few packets can be actually encoded (low coding opportunities) at the intermediate nodes. Hence the throughput benefits get nullified by the coding overheads, causing an initial degradation of performance.

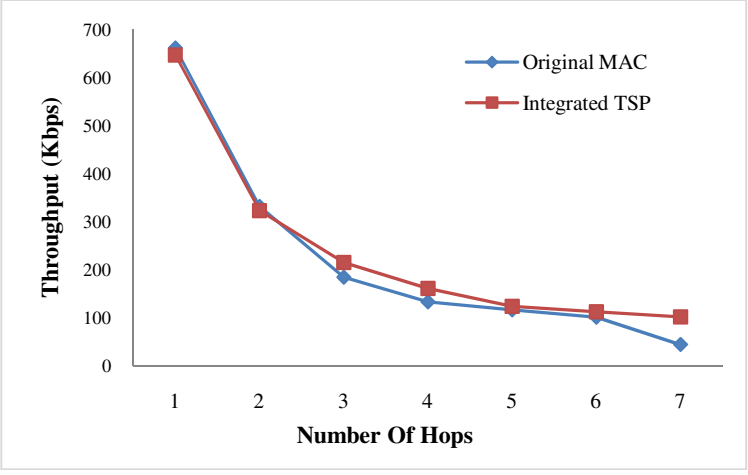


Fig. 7. Average Throughput VS Number of Hops

The effectiveness of the proposed DACK-XOR is realized as the available bandwidth of the network increases. As soon as the link capacity increases, the TCP sender tends to fill it by injecting more and more packets into the network thus increasing the overall coding probability. This implies that the number of saved transmissions increase and, consequently, a further portion of the link capacity results now available to the TCP sender.

6.4 RTS Transmissions

We measure the intra-flow contention by the ratio of RTS Transmissions to Packets Generated. The values shown in Table 3 depict that there is a decrease of RTS transmissions. Our DACK-XOR network coding scheme combines the DATA and ACK packets and sends them in one transmission, thereby reducing RTS ratio. The table shows the relative comparison of the percentage of RTS transmissions when compared with the number of actual packets generated during our simulation.

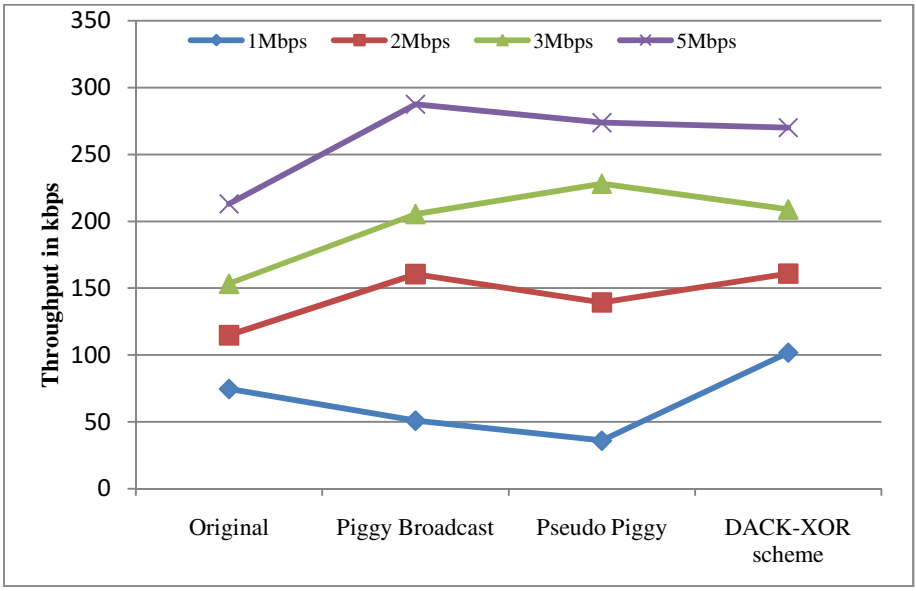


Fig. 8. Average Throughput for Different Variants for Different Data Rates

Table 3. Number of bytes generated and RTS count for Different Variants

Description	Generated Bytes	RTS Transmission	Ratio of RTS TX to Gen. Bytes
Original 802.11 MAC	4360000	85758	1.967%
Piggy code with Broadcast	6208000	111421	1.79%
Piggy code with Pseudo Broadcast	5308000	99256	1.87%
Integrated TSP	6015000	96648	1.61%

7 Conclusion and Future Work

This work presents an improvised network coding scheme that addresses the issue of intra-flow contentions. We reduce the number of packet transmissions by combining DATA and ACK packets travelling in opposite directions, performing a pseudo broadcast, and at the same time, take the stress off the decoding buffers, with the help of the cross-layer, rate based TCP which prevents packet clogging- A condition that is most likely to occur when the same node repeatedly wins medium contention. Although we have optimized the network coding buffer by retaining only the last packet delivered, achieving a decoding probability of 1 is not possible. On an average, for every 8000 TCP packets sent, we have 1 decoding failure. Moreover, we have realized our design by making changes to the TCP layer in addition to those at the MAC layer. Our future work would be to achieve buffer optimizations for network coding at MAC level alone instead of using a TCP variant to indirectly improve the efficiency of network coding.

References

1. Nage, T., et al.: TCP-Aware Network Coding with Opportunistic Scheduling in Wireless Mobile Ad Hoc Networks. In: Consumer Communications and Networking Conference (CCNC). IEEE (2010)
2. Siva RamMurthy, C., Manoj, B.S.: Ad hoc Wireless Networks: Architecture and Proto-cols
3. Network World – Network coding: networking's next revolution? `http://www.networkworld.com/news/2007/121007-network-coding.html`
4. Ahlswede, R., Cai, N., Li, S.-Y.R.: Network Information Flow. IEEE Trans.Inform. Theory (2000)
5. Deb, S., Effros, M., Ho, T., Karger, D.R., Koetter, R., Lun, D.S., Medard, M., Ratna-kar, N.: Network coding for wireless applications: A brief tutorial. IWWAN (2005)
6. Chen, C.-C., Oh, S.Y., Tao, P., Gerla, M., Sanadidi, M.Y.: Pipeline Net-work Coding for Multicast Streams (Invited Paper). Proc. of the ICMU (2010)
7. Wu, Y., Chou, P.A., Kung, S.Y.: Information Exchange in Wireless Networks with Network Coding and Physical-layer Broadcast, MSR-TR (2004)
8. Katti, S., Rahul, H., et al.: Xors in the air: Practical wireless network coding. IEEE/ACM Trans. Netw (June 2008)
9. Sundararajan, J., Shah, D., Medard, M., Mitzenmacher, M., Barros, J.: Network Coding Meets TCP. IEEE INFOCOM (April 2009)
10. Qin, C., Xian, Y., Gray, C.: MIX: Integration of Intra-Flow and Inter-Flow Wireless Network Coding. In: Proc. of IEEE SECON Work-Shops (2008)
11. Scalia, L., Soldo, F., Gerla, M.: PiggyCode: a MAC layer network coding scheme to improve TCP performance over wireless networks. IEEE GLOBECOM (2007)
12. Chen, C.-C., Chen, C., Oh, S.Y., Gerla, M., Sanadidi, M.Y.: ComboCod-ing: Combined Intra/Inter-Flow Network Coding for TCP over Disruptive MANETs. J. Adv.Res. 2(3) (2011)

Network Security Situation Prediction: A Review and Discussion

Yu-Beng Leau* and Selvakumar Manickam

National Advanced IPv6 Centre (NAv6), Universiti Sains Malaysia
11800, Bayan Lepas, Penang, Malaysia
{beng,selva}@nav6.usm.my
http://www.springer.com/lncs

Abstract. The rapid development of information technology exposes peoples life and work to the network. While people are enjoying in sharing their resources in the convenient condition, network security issues have emerged. Instead of considering security of single device in the network, researchers have shown an increased interest to grasp the overall network situation as a big picture in order to create situation awareness which consists of event detection, situation assessment and situation prediction. As the highest level in situation awareness, Network Security Situation Prediction makes quantitative prediction of incoming network security posture based on historical and present security situation information. The purpose is to provide an informational reference to network managers for helping them in formulating and implementing timely preventive measures before the network is under attack. In this paper, the authors group the existing network security situation prediction mechanisms into three major categories and review each model in the aspect of its strengths and limitations. The authors conclude that adaptive Grey Verhulst is more suitable to be used in predicting incoming network security situation.

Keywords: Network Security Situation Prediction, Machine Learning, Markov Model, Grey Theory.

1 Background

Recent years, network penetrates into our life and work with providing convenience services such as information sharing, resource accessing and etc. However, new security challenges are emerging while people are sharing their resources in this convenient condition. An investigation report stated that there have 63,437 security incidents recorded by 50 organizations from various industries around the world in 2013 [1] and it has reached an alarming level and begins to threaten the internet users in their daily activities. In fact, literatures showed that Intrusion Detection System (IDS) and Intrusion Prevention System (IPS) have

* National Advanced IPv6 Centre (NAv6), Universiti Sains Malaysia, 11800, Bayan Lepas, Penang, Malaysia.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 424–435, 2015.
DOI: 10.1007/978-3-662-46742-8_39

become preferred security defence mechanisms in many companies. They use IDS to analyze the audit log and suspicious packets and IPS to take the appropriate response against the attacks. Unfortunately, in a study done by University of South Wales in 2013 on nine big-brand IPS systems, they found that seven out of them were failed to detect and block 34%-49% of attacks that target vulnerabilities in web-based application [2]. All the remedies only can be taken after the suspicious attempts being detected. It directs the company network to a risky state where they are unable to predict the future security situation of the network.

2 Situation Prediction in Network Security Situation Awareness

Network Security Situation Awareness (NSSA) is first introduced by Tim Bass which adapted from the concept of Situation Awareness introduced by Endsley in 1988 [3]. He claimed that the next generation cyberspace IDS should fuse the network data from multiple or heterogeneous distributed sensors which located at network border and interpret them by decision maker in order to frame NSSA [4]. Basically, NSSA can be divided into three stages which are event detection, current situation assessment and future situation prediction. Event Detection identifies the abnormal and malicious activity in the network and translates them into logical format. Current Situation Assessment is a process to evaluate the security situation of the entire network by using the information obtained from previous stage. Last stage is Future Situation Prediction is aimed to forecast the future network security tendency according to the current and historical network security situation status. It purposes to provide an informational reference to network managers which able to help them in formulating and implementing timely preventive measures before the network is under attack. This changing of network security management from passive to active is tends to reduce the potential harm caused by attacks by improving the emergency response capacity. At present, the existing network security prediction techniques can be grouped into three main categories which are Machine Learning, Markov Model and Grey Theory. In this paper, the authors review each of them in terms of their strengths and limitations.

2.1 Based on Machine Learning

Machine learning is a scientific discipline that applies a computer-based resource to implement learning algorithms which enable computer to recognize pattern automatically and making decision intelligently based on training sample data without explicitly programmed. It detects the patterns in data and adjusts algorithm actions accordingly. The self-learning and adjusting features in its application has been extended to predict network security situation recent years. Researchers begin to apply the concept of Machine Learning such as Neural Network and Support Vector Machine to design their security situation prediction models.

Neural Network. Neural networks represent a kind of computing that simulates the way that the brain performs computations. It is a computer program that attempt to recognize underlying relationship in a set of data by using a process that mimics the way the human brain operates and learn from these identified relationships to predict the future patterns. The field of neural networks was pioneered by McCulloch and Pitts when they introduced the first neural network computing model in 1943 [5]. Basically, a Neural Network consists of four main parts which are processing units, weighted interconnection among processing units, an activation function in input signals to produce or activate the output signal and a learning function that makes the weights for a given input/output pair adjustable. There are three layers in a Neural Network which are input layer, hidden layer and output layer. There are built up of a number of interconnected nodes which contain an activation function. Various patterns from heterogeneous observation are presented to the network through the input layer which communicates to one or more nodes in hidden layer. The actual processing is done in hidden layer with its learning function on weighted connections. This layer links to an output layer where the evaluation of values of output variables allows user to be aware of a situation. The advantages of Neural Network is its high degree of fault tolerance with its variable connection weights matrix between the neurons, ability in self-learning and self-organizing and strong nonlinear mapping and generalization in complex system [6,7]. With these significant features in Neural Network, it has become an ideal tool used to predict the network security situation recently [8].

In 2008, Zongming et. al. proposed a network security situation model by using Back Propagation (BP) Neural Network [9]. This variant of neural network is a feed-forward network with multi-layers and spreading error from back to front while adjusting the parameters [10]. The network requires iterated modification of weight value and threshold value until the convergence error met. To optimize the BP Neural Network, the proposed model applies Particle Swarm Optimization (PSO) to reach global optimization of its weight value and threshold value. The flexibility in modifying these parameters is able to enhance the precision rate but the training to the sample is required. Consequently, over adaptation may occur where the network remembers the trained sample but lacks generalization ability of new sample.

Few years later, a network security situation prediction method based on dynamic BP Neural Network with covariance in 2011 [11] and 2012 [12,13]. In their model, the impact of sample covariance and noise on the network training is considered and the traditional function of error is replaced by the maximum likelihood error function. Through the error analysis, the predicted error value will be obtained and feedback to the prediction model as the training signal for the situation index weights adjustment. The improvement of precision could be achieved by taking previous prediction error into account for next prediction process. Unfortunately, training is required and the success of application is highly depends on the quality of training sample. Furthermore, it is not suitable for small scaled data because less input information will slower the convergence.

Aside from BP Neural Network, there is a kind of neural network called Wavelet Neural Network also had been applied to predict the network security situation [14]. Wavelet Neural Network was firstly introduced by Zhang Qinghua in 1992 [15]. The network uses wavelet analysis as the front-end processor to process the data and supplies the input vector to the neural network in loose combination mode and then replaces this function with activation function of hidden nodes directly in compact combination. With optimizing the model by improved genetic algorithm to encode, compute the fitness and operate the genetic before forecasting the non-linear time series of network security situation, this proposed model has high convergence rate, strong fault tolerance capacity and excellent in self-learning and adaption. But the model limited to certain application which the function to be chosen due to its suitability of the architecture. Moreover, the training is needed to gain the architecture parameters, weight value, dilation coefficient and translation coefficient.

Due to the uncertainly changing security environment in a complex system, a set of more flexible adaptive learning neurons has been introduced into the neural network concept for security situation prediction [6]. The neurons are able to work well in three basic functions in the model. There are positive dissemination of information, inverse dissemination of error and adjusting their parameters by themselves. The model has good performance where the functions can be done in parallel with these independent neurons and only the parameters of attacked neuron need to be revised instead of all neurons in the network. Nevertheless, to establish self-learning neuron is a difficult task and adjusting the steepness of activation function of situation prediction neuron is also challenging.

Notably, all the neural network prediction models are require training algorithm to generate the trained sample for prediction the incoming security situation. In order to simplify the training process, a network security situation prediction method based on small-world Echo State Networks has been suggested in 2013 [16]. Generally, Echo State Networks was initiated by H. Jaeger of Germany Jacobs University in 2001 [17]. A dynamic reservoir which is constructed by numerous neurons was allocated in the hidden layer of the network. The training algorithm in the model is simple and it is suitable to be used in approximating nonlinear dynamic system. However, adequate training data as input is needed and the output weight in the model is difficult to be prepared.

Support Vector Machine. Support Vector Machine (SVM) was proposed by Vapnik et al. in 1992 [18] and widely used in classification and regression. It maps the input space vector to a high-dimensional feature space. In other words, the nonlinear regression problem in low dimensional feature space has been converted to linear regression in high-dimensional feature space.

Compared with Neural Network, SVM has quick convergence rate and strong ability to resists a fitting [7]. The concept has been adopted by Xiaorong Cheng and the team to design their network security situation prediction model in 2012 [19]. In order to avoid the transformation of high-dimensional hyperplane in SVM, the kernel functions were introduced. With given training set, the model chooses the appropriate precision parameter and kernel function. Then

using these parameters, the model tries to solve the optimization problem. The strength of this SVM prediction model is relatively easy to train and seek the overall optimally. Furthermore, the model is also able to control the complexity and error of the classification. Unfortunately, adequate training sets are needed to gain the optimized parameters. In addition, the parameters of punishment factor and kernel functions are hardly to be determined in this proposed model. Blindness of parameters selection of SVM training process is also a main consideration when using it in prediction [7].

2.2 Based on Markov Model

Markov Model is a stochastic approach to describe the transitions from one state to another with its probabilities associated with various state-changes [20]. In a Markov Model, it consists of a list of the possible states of that system, the possible transition paths between those states and the probabilities of those transitions. The applications for machine learning is broad which cover from speech and handwriting recognition, medical diagnosis, credit card fraud detection as well as stock market and currency rate prediction.

In 2010, Dapeng Man et. al. has proposed a combination model for security situation prediction which combined with Autoregressive-Moving-Average (ARMA) model and Markov model [21]. They claimed that precision can be improved with fully utilizing each prediction method. In their model, the previous data will be used as input for both models separately to obtain its prediction value. The previous error of each model will be relatively calculated in order to obtain the weight for prediction model. Then the prediction value from combined model is expressed as

$$\sum_{n=1}^{N} w_n f_{nt} = w_1 f_{1t} + w_2 f_{2t} + ... + w_n f_{nt} \tag{1}$$

where f_{nt} represents the combined predicted value of the n kind of prediction models at time t and w_n represents the weight of the n^{th} prediction model. Compared with single prediction model, the proposed model has achieved higher precision rate. But the difficulties in selecting appropriate parameters in each model will affect the model performance. Moreover, their claim is arguable while high precision rate is not guaranteed if the combination is more than two prediction methods or from other different methods. In addition, although taking previous error into next prediction process is encouraged, but it is unable to improve significantly the precision in the next prediction.

Instead of designing combination model, GuangCai Kuang et. al. has presented a fuzzy prediction method of network security situation based on Markov [20]. In the proposed model, the correlation of network security is represented as Markov state transition matrix. With the current status of security situation, the improved linear weighted Zadeh formula is applied and membership effect matrix is used to calculate security situation values of predicted network. This

model has high precision rate with taking into account the impacts and influences of vulnerability but the training dataset is needed in constructing reference Markov matrix. Besides this, the assessment and prediction value in this model are keep increasing and they will reach the maximum point if the total length of time is big. Furthermore, determining the probability of all possible states and transitions subjectively especially in a complex network is a challenge.

A similar prediction model as previous work which based on fuzzy Markov Chain also been presented in 2014 [22]. The model utilizes historical data of safe behavior with the level of threat to forecast the network situation. After calculating threat value of particular period by using text categorization and threat level division technique, a set of fuzzy states with its membership degree has been built. Then the fuzzy state transition frequency has been calculated. Although the researcher claimed there is a good predictive performance for this proposed method, but assigning threat values according to five levels of attack without considering the various t level of the attack towards different devices is inappropriate. Additionally, the training is needed to derive the most suitable membership degree equation in different system. Another limitation of the model is it unable to handle unknown threat since it is based on the classification of known threats.

2.3 Based on Grey Theory

Grey Theory was initiated by Deng Ju-Long in 1982 in China [23]. It is used to predict from the grey system which lack of information. In this context, grey means poor, small-scaled, incomplete, uncertain data [24], which means the system information is partly know and partly unknown. With this noticeable feature, grey theory is widely used in various system to provide better prediction in short-term forecasting by capturing dominance at small sample. Basically, grey theory is applied to build a dynamic model with a group of differential equations [25,26], which is called Grey Model (GM) by using the least 4 data to replace difference modeling in vast quantities of data [24]. It does not need many data on the particular probability distribution. As superiority to statistical models, the model is able to prevail over the weakness of probability [27] and discovered the relationship among the limited and confused data [28]. It utilizes the sequence generated by Accumulated Generating Operation (AGO) to weaken the randomness of the original sequence [28]. This makes the process to find out the variation regularity in the sequence easier and use this regularity to forecast [29]

Grey Model (1,1). In grey theory, a First-order One-variable grey model (GM(1,1)) is the most widely used grey model. The modeling algorithm is described below:

Step I: Assume that the original raw data series $x^{(0)}$ with n samples is expressed as:

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), ..., x^{(0)}(n)\}, n \geqslant 4 \tag{2}$$

where $X^{(0)}$ is a non-negative sequence and n is the sample size of the data.

Step II: A new series $X^{(1)}$ is generated by applying Accumulating Generation Operation (AGO)

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), ..., x^{(1)}(n)\}, n \geqslant 4 \tag{3}$$

where

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, 3, ..., n \tag{4}$$

Step III: A series $Z^{(1)}$ is generated by applying the MEAN operation to $X^{(1)}$

$$Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), ..., z^{(1)}(n)\}, \tag{5}$$

where $z^{(1)(k)}$ is the mean value of adjacent data such as

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1), k = 2, 3, ..., n \tag{6}$$

The least square estimate sequence of the grey difference equation of GM(1,1) is defined as follows [24]:

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{7}$$

and its whitening equation is

$$\frac{\partial x^{(1)}(k)}{\partial t} + ax^{(1)}(k) = b \tag{8}$$

in which a and b are the interim parameters. The parameter matrixes are

$$\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y \tag{9}$$

where $$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \tag{10}$$ $$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{11}$$

According to equation (7), the solution of $x^{(1)}(t)$ at time k is

$$x_p{}^{(0)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak} + \frac{b}{a} \tag{12}$$

Step IV: To obtained the predicted value of the primitive data at time *(k+1)*, the Inverse Accumulating Generation Operation (IAGO) is used to establish the following grey model.

$$x_p{}^{(0)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right] e^{-ak}(1 - e^{a}) \tag{13}$$

In 2006, Lai Jibao claimed that the grey theory is suitable for building the prediction model and able to ensure its prediction. Hence, he first applied grey

theory in constructing the network security situation prediction model based on grey theory which is built by past and current network security situation [30]. After three years, a dynamic prediction model has been presented by Fengli Zhang. He proposed to estimate the overall network security situation by applying grey forecast model. By using classical GM(1,1), the models calculates the future value of situation and computes the absolute difference and relative difference between the forecast and real values. If the square sum of remnant difference is very large which represent less precision in forecast value, the model will modify the value by referring to the remnant difference [31]. Without any training on the input data, both of the proposed GM(1,1) models are working well with linear situation but not suitable for non-stationary random sequence.

To overcome the limitation of GM(1,1) model, Li Juan has proposed a hybrid prediction model which can be used in high fluctuation system in 2009. They combined unbiased Grey Theory and Markov forecasting theory to predict network security dynamic situation. Firstly, unbiased GM(1,1) model is used to obtain the forecasting value. Then by adjusting the result of unbiased GM(1,1), Markov model is used to determine the state, calculate the evaluation value and construct the transition probability matrix. The median of gray interval is used as forewarning value after the tendency of the system condition developing has been decided [32]. Unfortunately, there is a challenge to determine subjectively the probability of state. On the other side, Liu Nian also suggested another hybrid model to forecast the network security situation in the same year. He applied the combination of Grey Theory and Markov Theory to predict the incoming network security situation which is high randomness, uncertainty and fluctuation. During the process, classical GM(1,1) model is used to predict the situation data and find out its changing trend. Then, Markov Theory is employed to modify the error in the model in order to improve the prediction accuracy of network security situation changing trend [33]. Both of these hybrid models require training value to divide the original data into different states and modify the error in the model respectively. In order to improve the precision rate of GM(1,1) model, Rongzhen Fan has included a three-phase grey residual error correction model in GM(1,1) model. The proposed correction model used concept of GM(1,1) to calculate difference of the predict value and original accumulate value in particular time-frame [34]. The good thing of his work is taking into consideration the residual error for predicted value. But in the same time, high processing power is needed for recurrent residual error correction process.

In fact, the conventional GM(1,1) model is only suitable for the prediction with strong exponential law and it only able to depict the monotone variation [35]. The model is imperfect when the series increases in the curve with S type or the increment of series is in the saturation stage [36]. These limitations have sparked the argument on suitability of classical GM(1,1) to be used to predict accurately the future status of network security situation which has variation characteristics [16]. Instead of depending on linear differential equation in GM(1,1), a variant of grey model called Grey Verhulst model has been used in predicting the non-linear series.

Grey Verhulst. The Verhulst model was first introduced by a German biologist, Pierre Franois Verhulst in 1837 to describe the increasing process like S-curve which has a saturation region, namely the process is increasing slowly at initial stage, then speed-up and finally grow slowly or stop growing [37]. Same as GM(1,1), Grey Verhulst model has superiority in small sample. The Grey Verhulst model can be defined as follows [38]):

The equations (1) to (5) in GM(1,1) are same in Grey Verhulst Model. To predict the S type curve, the model is applying a non-linear difference equation as below

$$x^{(0)}(k) + az^{(1)}(k) = b(z^{(1)}(k))^2 \tag{14}$$

and its whitening equation is

$$\frac{\partial x^{(1)}(k)}{\partial t} + ax^{(1)}(k) = b(x^{(1)}(k))^2 \tag{15}$$

in which a is defined as the development coefficient and b is grey input. As equation (9), the parameter matrixes B and Y are

$$B = \begin{bmatrix} -z^{(1)}(2) & (z^1(2))^2 \\ -z^{(1)}(3) & (z^1(3))^2 \\ \vdots & \vdots \\ -z^{(1)}(n) & (z^1(n))^2 \end{bmatrix} \tag{16}$$

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{17}$$

By calculating equation (15), the solution of $x^{(1)}(t)$ at time k is

$$x_p{}^{(1)}(k+1) = \frac{ax^{(0)}(1)}{bx^0(1) + (a - bx^{(0)}(1))e^{ak}} \tag{18}$$

In the equation (19), $x^{(0)}(1) = x^{(1)}(1)$. It is assumed that the n-dimension data sequence is selected to fit the model. The fitted model can be used to predict the future value as

$$x_p{}^{(0)}(k+1) = x_p{}^{(1)}(k+1) - x_p{}^{(1)}(k), k \geq n \tag{19}$$

where

$$x_p{}^{(0)}(1), x_p{}^{(0)}(2), x_p{}^{(0)}(3), ..., x_p{}^{(0)}(n) \tag{20}$$

are called Grey Verhulst fitted sequence, while

$$x_p{}^{(0)}(n+1), x_p{}^{(0)}(n+2), x_p{}^{(0)}(n+3), ..., x_p{}^{(0)}(n+t) \tag{21}$$

are called Grey Verhulst predicted values.

In Grey Verhulst model, a and b are the key parameters to guarantee the precision of the model. The values of them can be obtained by applying least square method into the generation sequences $Z^{(1)}$ as equation (5). This feature is only allow the Grey Verhulst model to generate appropriate parameters in the small time interval and the AGO curve varies smoothly.

In 2010, Hu Wei found out that the generated sequence will make the prediction generate the advance or delay error which will depress the model precision. Thus, he first adopts the adaptive determination of the grey parameters to Grey Verhulst model to guarantee the precision. He made the assumption of the function of AGO curve as

$$X^{(1)}(t) = \frac{1}{\alpha e^{\beta t}} \tag{22}$$

and calculating the area below the curve by integration method [29]. His model is only applicable to estimate the curve trend accurately if the S-type risk growing situation is matches to the AGO curve function. Furthermore, the model is also limited to a single-peak situation variation instead of multiple-peaks.

3 Conclusion

In this study, the authors conclude that each group of prediction mechanism has its strengths and limitations. Machine Learning is excellent in self-learning and self-adaption and able to provide high convergence rate as well as strong fault tolerance capacity. But the adequate training data is required to gain the parameters and it is difficult to establish the neurons with self-learning and adaption capabilities. For Markov Model, although it is able to perform in various time series prediction, a set of training data is still needed. Moreover, all possible states and its transitions especially in a complex network are hardly to identify. Meanwhile, Grey Theory can provide better prediction in short-term forecasting with small sample data without any training required. Even though GM(1,1) is only limited to linear time series prediction and the generation sequence with mean is only suitable for small time interval, but adaptive Grey Verhulst is surpassing with its adjustable generation sequence and non-linear S-curve time series prediction. Considering the chronology of intrusion attack, Grey Verhulst is more suitable to predict the incoming network security situation with its remarkable features.

Acknowledgement. The authors would like to thank the National Advanced IPv6 Centre (NAv6), Universiti Sains Malaysia for supporting this research project.

References

1. 2014 Data Breach Investigations Report. pp. 1-60. United States: Verizon Enterprise (2014)
2. Xynos, K., Sutherland, L., Blyth, A.: Effectiveness of Blocking Evasions in Intrusion Prevention System. University of South Wales, pp. 1–6 (2013)
3. Endsley, M.R.: Situation Awareness Global Assessment Technique (SAGAT). In: 1988 National Aerospace and Electronics Conference, pp. 789–795. IEEE Press (1998)
4. Bass, T.: Intrusion Detection Systems and Multisensor Data Fusion. Communications of the ACM 43(4), 99–105 (2000)

5. McCulloch, W.S., Pitts, W.: A Logical Calculus of the Ideas Immanent in Nervous Activity. The Bulletin of Mathematical Biophysics 5(4), 115–133 (1943)
6. Li, J., Dong, C.: Research on Network Security Situation Prediction-Oriented Adaptive Learning Neuron. In: Second International Conference on Networks Security Wireless Communications and Trusted Computing (NSWCTC), pp. 483–485. IEEE Press (2010)
7. Wei, X., Jiang, X.: Comprehensive Analysis of Network Security Situational Awareness Methods and Models. In: 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), pp. 176–179. IEEE Press (2013)
8. Xi, R., Jin, S., Yun, X., Zhang, Y.: CNSSA: A Comprehensive Network Security Situation Awareness System. In: IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 482–487. IEEE Press (2011)
9. Lin, Z., Chen, G., Guo, W., Liu, Y.H.: PSO-BPNN-based Prediction of Network Security Situation. In: 3rd International Conference on Innovative Computing Information and Control, pp. 37–41. IEEE Press (2008)
10. Zhang, Y., Jin, S., Cui, X., Yin, X., Pang, Y.: Network Security Situation Prediction Based on BP and RBF Neural Network. In: Yuan, Y., Wu, X., Lu, Y. (eds.) ISCTCS 2012. CCIS, vol. 320, pp. 659–665. Springer, Heidelberg (2013)
11. Tang, C., Xie, Y., Qiang, B., Wang, X., Zhang, R.: Security Situation Prediction Based on Dynamic BP Neural with Covariance. Procedia Engineering 15, 3313–3317 (2011)
12. Tang, C., Wang, X., Zhang, R., Xie, Y.: Modeling and Analysis of Network Security Situation Prediction Based on Covariance Likelihood Neural. In: Huang, D.-S., Gan, Y., Premaratne, P., Han, K. (eds.) ICIC 2011. LNCS, vol. 6840, pp. 71–78. Springer, Heidelberg (2012)
13. Zheng, R., Zhang, D., Wu, Q., Zhang, M., Yang, C.: A Strategy of Network Security Situation Autonomic Awareness. In: Lei, J., Wang, F.L., Li, M., Luo, Y. (eds.) NCIS 2012. CCIS, vol. 345, pp. 632–639. Springer, Heidelberg (2012)
14. Lai, J.B., Wang, H.Q., Liu, X.W., Liang, Y., Zheng, R.J., Zhao, G.S.: WNN-based Network Security Situation Quantitative Prediction Method and Its Optimization. Journal of Computer Science and Technology 23(2), 222–230 (2008)
15. Zhang, Q., Benveniste, A.: Wavelet Networks. IEEE Transactions on Neural Networks 3(6), 889–898 (1992)
16. Chen, F., Shen, Y., Zhang, G., Liu, X.: The Network Security Situation Predicting Technology Based on the Small-world Echo State Network. In: 4th IEEE International Conference on Software Engineering and Service Science, pp. 377–380. IEEE Press (2013)
17. Jaeger, H.: Tutorial on Training Recurrent Neural Networks, Covering BPPT, RTRL, EKF and the" Echo State Network" Approach. GMD-Forschungszentrum Informationstechnik (2002)
18. Cortes, C., Vapnik, V.: Support-vector Networks. Machine Learning 20(3), 273–297 (1995)
19. Cheng, X., Lang, S.: Research on Network Security Situation Assessment and Prediction. In: Fourth International Conference on Computational and Information Sciences, pp. 864–867. IEEE Press (2012)
20. GuangCai, K., XiaoFeng, W., LiRu, Y.: A Fuzzy Forecast Method for Network Security Situation Based on Markov. In: International Conference on Computer Science and Information Processing, pp. 785–789 (2012)

21. Man, D., Wang, Y., Wu, Y., Wang, W.: A Combined Prediction Method for Network Security Situation. In: International Conference on Computational Intelligence and Software Engineering, pp. 1–4. IEEE Press (2010)
22. Wang, Y., Li, W., Liu, Y.: A Forecast Method for Network Security Situation Based on Fuzzy Markov Chain. In: Huang, Y.-M., Chao, H.-C., Deng, D.-J. (eds.) Advanced Technologies, Embedded and Multimedia for Human-centric Computing. LNEE, vol. 260, pp. 953–962. Springer, Heidelberg (2014)
23. Ju Long, D.: Control Problems of Grey Systems. Systems & Control Letters 1(5), 288–294 (1982)
24. Deng, J.-L.: Introduction to Grey System Theory. The Journal of Grey System 1(1), 1–24 (1989)
25. Deng, J.-L.: Modelling of the GM model of Grey Systems, pp. 40–53 (1988)
26. Deng, J.-L.: Properties of the Grey Forecasting Model of GM(1,1). Grey System, pp. 79–90 (1988)
27. Liu, S., Forrest, J., Yang, Y.: A Brief Introduction to Grey Systems Theory. Grey Systems: Theory and Application 2(2), 89–104 (2012)
28. Kordnoori, S., Mostafaei, H., Kordnoori, S.: The Application of Fourier Residual Grey Verhulst and Grey Markov Model in Analyzing the Global ICT Development. Hyperion Economic Journal 2(1), 50–60 (2014)
29. Hu, W., Li, J.-H., Chen, X.-Z., Jiang, X.-H.: Network Security Situation Prediction Based on Improved Adaptive Grey Verhulst Model. Journal of Shanghai Jiaotong University (Science) 15, 408–413 (2010)
30. Jibao, L., Huiqiang, W., Liang, Z.: Study of Network Security Situation Awareness Model Based on Simple Additive Weight and Grey Theory. In: International Conference on Computational Intelligence and Security, pp. 1545–1548. IEEE Press (2006)
31. Zhang, F., Wang, J., Qin, Z.: Using Gray Model for the Evaluation Index and Forecast of Network Security Situation. In: International Conference on Communications, Circuits and Systems, pp. 309–313. IEEE Press (2009)
32. Juan, L., Tao, L., Gang, L.: A Network Security Dynamic Situation Forecasting Method. In: International Forum on Information Technology and Applications, pp. 115–118. IEEE Press (2009)
33. Diangang, W., Xuemei, H., Sunjun, L., Kui, Z.: Research on Network Security Situation Awareness Technology Based on Artificial Immunity System. In: International Forum on Information Technology and Applications, pp. 472–475. IEEE Press (2009)
34. Rongzhen, F., Mingkuai, Z.: Network Security Awareness and Tracking Method by GT. Journal of Computational Information Systems 9(3), 1043–1050 (2013)
35. Liu, S.F., Lin, Y.: An Introduction to Grey Systems Theory. IIGSS Academic Publisher, Grove City (1998)
36. Guo, Z., Song, X., Ye, J.: A Verhulst Model on Time Series Error Corrected for Port Throughput Forecasting. Journal of the Eastern Asia society for Transportation studies 6, 881–891 (2005)
37. Wang, Z., Dang, Y., Wang, Y.: A New Grey Verhulst Model and Its Application. In: International Conference on Grey Systems and Intelligent Services, pp. 571–574. IEEE Press (2007)
38. Wen, K.-L., Huang, Y.-F.: The Development of Grey Verhulst Toolbox and the Analysis of Population Saturation State in Taiwan-Fukien. In: International Conference on Systems, Man and Cybernetics, pp. 5007–5012. IEEE Press (2004)

Cloud and Parallel Computing

Lightweight Virtualization in Cloud Computing for Research

Muhamad Fitra Kacamarga, Bens Pardamean, and Hari Wijaya

[1] Bioinformatics & Data Science Research Center, Bina Nusantara University
Jakarta, Indonesia
bpardamean@binus.edu

Abstract. With the advancement of information technology and the wide adoption of the Internet, cloud computing has become one of the choices for researchers to develop their applications. Cloud computing has many advantages, particularly the ability to allocate on-demand resources without the need to build a specialized infrastructure or perform major maintenance. However, one of the problems faced by the researcher is the availability of computer tools to perform research. Docker is a lightweight virtualization for developers that can be used to build, ship, and run a range of distributed applications. This paper describes how Docker is deployed within a platform for bioinformatics computing.

Keywords: Virtualization, cloud computing, Docker, research.

1 Introduction

Computer programs are becoming more essential to many aspects of scientific research. The steps of the scientific process, from data collection, analysis, evaluations, and conclusions rely increasingly more on computational methods. With the advancement of computing technologies and the wide adoption of the Internet, computing resources have become cheaper, more powerful, and more prevalent, leading to a much higher availability than ever before [1]. One the impacts of these advancements is the presence of a new computing paradigm model called the cloud computing, in which resources (computing and storage) can be distributed as services that can be leased to customers through the Internet at an on-demand and as-needed term [2]. Within a cloud computing set-up, users utilize the resources provided by the infrastructure source as needed and pay only for those items [3].

Cloud computing has become a substantial tool of choice for researchers since it allows for the performance of complex computations and the exploration of new projects without the up-front investment in an expensive, customized infrastructure [4]. Additionally, researchers often do not have enough computational tools or time or expertise level to implement installations for a state-of-the-art data analysis application from scratch [5]. For instance, in the field of bioinformatics, state-of-the-art computational tools and algorithms for applications on biological, medical, and health data are essential for collection, sharing, and analysis [6]. Therefore, a complete computational environment that would assist collaboration between researchers is needed [7].

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 439–445, 2015.
DOI: 10.1007/978-3-662-46742-8_40

Docker is a lightweight virtualization for developers that can be used to build, ship, and run a range of distributed applications [8]. It can be used to build a *system image* that contains state-of-the-art data analysis application, which can then be shared with researchers involved in the project. This paper describes how Docker is deployed within a bioinformatics computing platform in cloud computing.

1.1 Cloud Computing

Cloud computing is a computing model that is an elastically scalable, virtualized system with the ability for rapid provision with minimal management effort over the Internet [9]. Google, Amazon, and Microsoft are market leaders for the cloud computing industry. The emergence of cloud computing has made several compelling features that makes it attractive to users:

1. No up-front investment: building a large-scale system would need a lot of investments in information technology infrastructure. The cloud computing uses a pay-as-go pricing model so users do not need to build the infrastructure themselves.
2. Elastic Infrastructure: the infrastructure can dynamically scale up or down based on request. Users can easily expand its infrastructure to a larger scale to handle rapid increase in service demands and shrink down when demand decreases.
3. Lower operating cost: resources in a cloud environment can be easily allocated and de-allocated, allowing users to manage resources more effectively and efficiently.
4. Easy Access: services provided in the cloud are generally web-based, rendering easy access through Internet connection. Users can also manage their resources using the cloud service provider's management console that can be accessed via Internet.
5. Reducing business risks and maintenance costs: cloud computing has access availability with high guaranteed uptime. By outsourcing the information technology infrastructure to the cloud, users can reduce its business. Users also do not need to hire staff and hardware maintenance since this task is covered by the cloud service provider.

Cloud computing provider offers three service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). This example of service models can be seen in Figure 1.

1. Infrastructure as a Service (IaaS) provides services for customer to access computing resources in a virtualized environment. These computing resources include computing unit, storage, network, and other fundamental computing resources.
2. Platform as a Service (PaaS) delivers services as a computing platform and includes operating system, programming language execution environment, and other tools for designing, building, and deploying the customer's application into the cloud infrastructure.
3. Software as a Service (SaaS) provides customers with access to the hosted applications on the cloud infrastructure that is managed by the vendor or the cloud service provider.

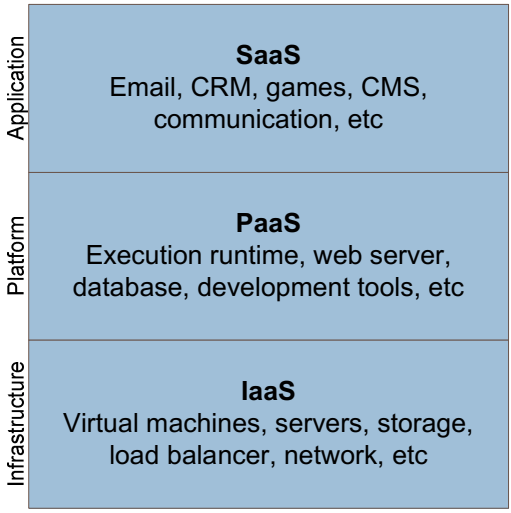


Fig. 1. An example of the cloud computing service models

Researchers can take advantage of the cloud computing service models depending on their research domain. There are many cloud computing providers from which to choose, such as Amazon Web Services, Google Cloud Platform, Microsoft Windows Azure, etc. Although each of them has unique components to provide different services, they also have several functionalities that are the same with one another. For example, Amazon EC2 provides the same service as Google Compute Engine, which provides virtual server instance. However, switching from one service provider to another is not easy in most cases due to *vendor lock-in* [10]. This is mainly due to dependencies and proprietary formats found within the underlying cloud infrastructures.

1.2 Docker

Docker is a lightweight virtualization based on Linux Containers (LXC) that can completely encapsulate an application and its dependencies within a virtual container [8]. This container can run on any Linux server enabling it to run on any premises (public cloud or private cloud) that has a Linux operating system. LXC is an operating system level virtualization technology that creates a sandbox virtual environment in Linux without the overhead of a virtual machine [5]. The overhead of a Docker's container is much smaller than that of a virtual machine because it replicates only the libraries and binaries of the application that is being virtualized [8]. Docker extends the LXC technology, leading to easier usage and simpler ways to perform versioning, distributing, and deploying. A Docker image can be transferred from one Docker host to another. Furthermore, it can be exported, archived, and run anytime in the future with the assurance of a similar computational environment. The comparison between a virtual machine and Docker architecture is shown in Figure 2.

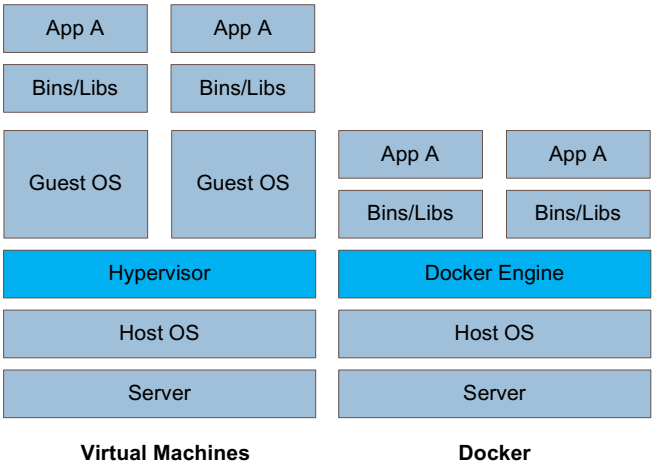


Fig. 2. Comparison between a virtual machine and the Docker architecture

Other than providing a virtualized and consistent computational environment, Docker provides many features that makes it an attractive sharing tool for research [5, 11]:

1. Docker images can be built by using text files (a Dockerfile) containing a set of instructions that commands Docker how to build its image. This approach allows Docker image to be versioned, shared, and re-built by others.
2. Docker has its own repository, similar to a Git repository. The Docker image can be easily shared with others via hosted repository, the Docker Hub.
3. The contents in a Docker container are restored to their original condition every time the container is launched. This approach makes a Docker container has a consistent computational environment.
4. Data, documentations, and files can be packaged within the Docker image, allowing for its use for sharing an entire computational experiment.
5. Docker has a large, active user base that provides a community-based information for troubleshooting.
6. Docker containers only use active resources, creating minimal overhead when running a Docker container.
7. Directories (folders) on the host's system can be easily mounted into a Docker container, rendering a seamless data sharing process between the host and the container.

There are two ways to build Docker containers: import from *tarball* file or import from a Dockerfile. Building a container from *tarball* file allows the researcher to create a complete container without performing installation steps. This method is similar with building a Virtual Machine from a *Snapshot* image file. On the other hand, building a container from a Dockerfile requires a series of installation commands (e.g., download application, libraries, etc.). Dockerfile is a text document that contains all commands of installation to build a Docker image. This method is more preferable when the researcher wants set up their computational environment within a

container. By installing from a Dockerfile, researchers can choose specific software for installation in the container.

2 Methods

This study describes a condition in which a researcher has existing virtual server instances but still requires a wide range of bioinformatics analysis. A computing platform can be used for data analysis ([R], Python, Perl) to perform a genome-wide association study (via PLINK). The system consists of a main installation script and a set of installation files. Users can edit the main installation script in order to select the bioinformatics software and create their own customized version. The objective for this system is to simplify the following processes: software selection, automatic new container establishment with the specified software, and deployment on the researcher's existing cloud virtual server instance.

3 Results

The bioinformatics computing platform was implemented using Docker. It is based on the Debian operating system. The researcher began platform deployment through a Linux virtual server instance. Then Docker was installed on its instance, followed by the installation package download. This installation package includes a Dockerfile, as well as other installation files. The researcher was also able to edit the Dockerfile to customize which bioinformatics tools had to be included in its container. In this Dockerfile, software packages were categorized based on their installation method. A set of bioinformatics tools including [R], Python, Perl, PLINK, and Cython were installed by the *apt-get* command. On the other hand, tools like Affymetrix Power Tools, Java, and Eigensoft were installed from the installation files provided within the installation package. When the Dockerfiles were executed, the Docker performed retrieval and installation of the selected software from the repositories and installation files, building the image into a fully functional bioinformatics computing container.

Researchers also had the ability to share the edited Dockerfile in order to distribute their own customized version of a bioinformatics platform with the selected software and data. Other researchers could replicate the customized bioinformatics platform by retrieving this edited Dockerfile then executing it in their cloud. Researchers could upload data from its local desktop computer or other source to the host. To share data from the host to a Docker container, copying the data to the mounted directory allows for the subsequent data retrieval from the Docker container. There was also the option to save a container with analysis result by exporting the container into *tarball* file, which could then be shared with other researchers for collaboration. Using the *tarball* file, the collaborators could create a whole identical container on their cloud then use new data with different parameter to generate new results. These steps are depicted in Figure 3.

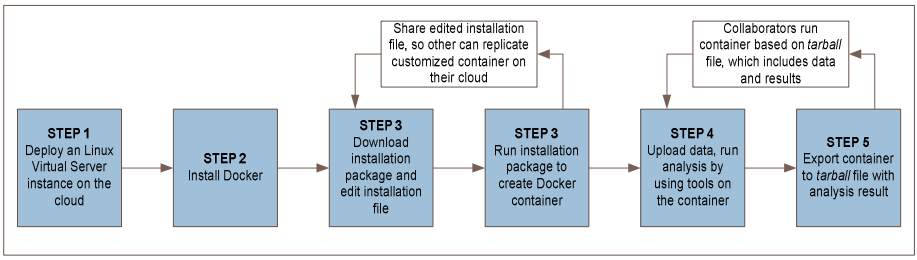


Fig. 3. Bioinformatics computing platform workflow

While research using the cloud computing environment supports reproducible research [12], many researchers worked locally, primarily with software installed on their local computer. Researchers would transform their work to become cloud-based only when collaborative tasks were performed or an increase in computational power was needed. Working locally allowed a researcher to exchange files and debugging faster. Docker is available on most major platform. Thus, a researcher could install Docker and build up the bioinformatics platform container on a local computer. The platform could also be used locally. The collaborators could also import a customized bioinformatics platform then run it locally for testing purpose.

4 Discussion

Several studies (Krampis [13] and Dudley & Butte [12]) have proposed ways to share computational environment for research. There are two dominant approaches: workflow software and virtual machines [11]. Workflow software provides solutions to standardize the creation, representation, and sharing of computational workflow that bind diverse software tools together into a single analysis [12]. This workflow software is often adopted by well-funded collaboration research through which they receive substantial support from the communities. Most workflow software has relatively low adoption due to proprietary formats and interfaces [12].

Virtual machine (VM) offers a more direct approach. The VM approach, operating system, software tools, and databases are packaged into a single digital image that is ready to be used. This approach is used by Krampis [13] and Dudley & Butte [12] to share virtual machine images that will run on the cloud as a platform for doing research. But, there are some drawbacks using VM images such as large file size, the need for system administration knowledge, and the difficulties to track versions [5]. The implementation of VM in cloud has a serious problem when users want to switch from one service provider to another due to vendor *lock-in* [10].

5 Conclusion

This study described how Docker was deployed for a bioinformatics computing platform in cloud computing. The motivation for this study was the primary concern on

the lack of time and expertise on a researcher's part to install and implement state-of-the-art data analyses application from the scratch. Our objective was to simplify the processes of software selection, automatic building of a new container with the specific softwares, and deployment on a researcher's existing cloud virtual server instance. A script (Dockerfile) that allows bioinformatics platform to be easily reproduced and updated was developed since the script provides an exact instruction on how the image was built. Furthermore, researchers also has the option to save the container with analysis result by exporting the container into a *tarball* file. This approach is an easy way for sharing a complete computational environment for researchers.

Acknowledgements. We would like to thank James W. Baurley, PhD from BioRealm Research for sharing the bioinformatics computing platform.

References

1. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges. J. internet Serv. Appl. 1, 7–18 (2010)
2. Marinescu, D.C.: Cloud Computing and Computer Clouds (2012)
3. Weinhardt, C., Anandasivam, W.A., Blau, B., Borissov, N., Meinl, T., Michalk, W.W., Stößer, J.: Cloud computing–a classification, business models, and research directions. Bus. Inf. Syst. Eng. 1, 391–399 (2009)
4. Rehr, J.J., Vila, F.D., Gardner, J.P., Svec, L., Prange, M.: Scientific computing in the cloud. Comput. Sci. Eng. 12, 34–43 (2010)
5. Chamberlain, R., Invenshure, L.L.C., Schommer, J.: Using Docker to support reproducible research (2014)
6. Kesh, S., Raghupathi, W.: Critical Issues in Bioinformatics and Computing. Perspect. Health Inf. Manag. 1, 9 (2004)
7. Hu, Y., Lu, F., Khan, I., Bai, G.: A Cloud Computing Solution for Sharing Healthcare Information (2012)
8. Hykes, S.: What is Docker?, `https://www.docker.com/whatisdocker/`
9. Padhy, R.P., Patra, M.R., Satapathy, S.C.: X-as-a-Service: Cloud Computing with Google App Engine, Amazon Web Services, Microsoft Azure and Force. Com. Int. J. Comput. Sci. Telecommun. 2, 8–16 (2011)
10. Kratzke, N.: A Lightweight Virtualization Cluster Reference Architecture Derived from Open Source PaaS Platforms. Open J. Mob. Comput. Cloud Comput. 1, 17–30 (2014)
11. Boettiger, C.: An introduction to Docker for reproducible research, with examples from the R environment. arXiv Prepr 0846 (2014)
12. Dudley, J.T., Butte, A.J.: Reproducible in silico research in the era of cloud computing. Nat. Biotechnol. 28, 1181 (2010)
13. Krampis, K., Booth, T., Chapman, B., Tiwari, B., Bicak, M., Field, D., Nelson, K.E.: Cloud BioLinux: pre-configured and on-demand bioinformatics computing for the genomics community. BMC Bioinformatics 13, 42 (2012)

A Cloud-Based Retail Management System

Adewole Adewumi, Stanley Ogbuchi, and Sanjay Misra

Department of Computer and Information Sciences, Covenant University, Ota, Nigeria
{wole.adewumi,stanley.ogbuchi,
sanjay.misra}@covenantuniversity.edu.ng

Abstract. Retail management systems have been deployed extensively as web applications and stand-alone systems. However, in order to maximize return on investment while also improving on retail business efficiency and performance, it is imperative to explore newer technologies that can be leveraged. Cloud computing shows great potential in this regard; and so it is our aim in this paper to develop a cloud-based retail management system. We realize this by first designing the framework of the system and then implementing it.

Keywords: Cloud computing, framework, retail management, point of sale.

1 Introduction

A retail management system is one that is used by retail-inclined organizations like super markets, hardware stores, shopping outlets, e-commerce stores, bookstores, pharmacies and a host of others [1]. Features such as Point of Sale (POS), inventory management, reports & feedbacks, employee management, customer management, and supply chain management are some key components of retail management systems [2]. These systems have mostly been developed and deployed as Web Applications and stand-alone systems [3]. Where they are implemented as stand-alone systems, they tend to be complicated and expensive to maintain by businesses that have little or no IT knowledge [4]. The amount and variety of hardware and software needed to run these systems can also be discouraging [5]. Retailers would require a whole team of experts to install, configure, test, run, secure, and update such systems. In order to maximize return on investment, newer technologies can be explored. Cloud computing in particular shows great potential in this regard [6].

Cloud computing represents a radical change in the way information technology (IT) services are offered to the public [6]. Its concept relies on sharing computing resources rather than owning local servers in order to handle applications [7]. Cloud solutions can be implemented under three major service models namely: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). They can be deployed as private clouds, public clouds, community clouds or hybrid clouds [8]. It is of utmost importance that cloud computing be greatly applied in the retail industries because it transverses borders and contributes greatly to any Nation's economy [9]. However, as retailers begin to gravitate towards the cloud-based approach to retail management, they will desire among other things – a platform independent solution and one that is easy to

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 446–456, 2015.
DOI: 10.1007/978-3-662-46742-8_41

customize to suit changing requirements. A framework that can serve as a guide for implementation and deployment is also desirable [10]. The rest of this paper is thus structured as follows: Section 2 reviews existing systems in this domain. In Section 3, we focus on the design of the proposed system. In Section 4, we implement the retail management solution called SkyRetail based on our design in Section 3. The key user interface modules are also highlighted and discussed in this section. Section 5 concludes the paper.

2 Review of Existing Systems

Microsoft Dynamics Retail Management System is an application from Microsoft that offers small and midsize retailers a complete point of sale (POS) solution that can be adapted to meet unique requirements [11]. It provides centralized control for multi-store retailers, and integrates with Microsoft Office system programs. It also offers benefits in; ease of use, automation, efficiency, flexible reporting and scalability. It can be deployed for any form of retail enterprise from pharmaceuticals to grocery stores. The system however is not cloud-based and also targets businesses that run Windows operating system – thus it is not cross-platform.

Lightspeed Cloud is another complete point of sale system [12]. It offers a range of complete retail solution for retailers. Among other things, it offers access to the system from both web and mobile browsers. It also handles inventory, customers/employee management and report generation. According to Lightspeed – the software company that developed the solution, their retail solution is: easy to use; continuously upgraded; cloud based; able to run on Web and mobile browsers; always available; easy to set up and install. The drawback of this system however is that it is not available outside of North America.

Epicor Cloud Retail Software is a SaaS retail solution from Epicor Solutions [13]. It serves small to mid-sized retailers who want to leverage their insufficient IT resources. Epicor delivers a model that significantly reduces capital investment, implementation challenges, and on-going requirements of managing IT. With Epicor, retailers are able to integrate their sales channels, order management, POS systems, inventory, and other operations for access to right information at the appropriate time. Epicor cloud retail software supports: merchandising, store operations, CRM, audit and operations management and planning. The system however is difficult to customize so as to suit a particular organization.

From the drawbacks identified in the existing systems, we intend to realize a retail management solution that is cloud-based; platform independent; able to run on Web and mobile browsers and also easy to customize by any retail organization. This is the motivation behind this paper.

3 Design of the Proposed System

Being software, we incorporate the retail management system into our cloud design framework by modeling it at the SaaS layer - above the PaaS and IaaS layer. The design framework of the proposed system uses a layered structure as depicted in Fig. 1.

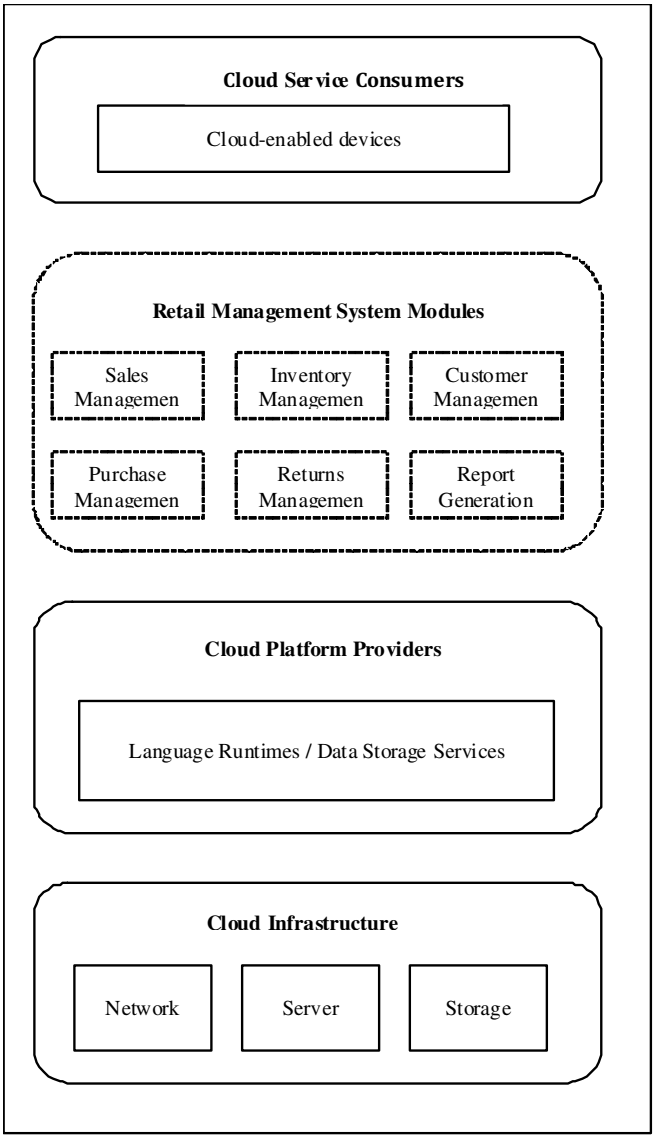


Fig. 1. Framework of the proposed retail management system

3.1 Cloud Service Consumers

This refers to devices that will be used to access and use the retail management system. Such devices should be cloud-enabled and include: laptops, tablet computers and smartphones.

3.2 Retail Management System Modules

The proposed retail management system (shown with dashed lines in Fig. 1) will have the following key features: inventory management that makes it easy to add and track retail items; customer management that helps to keep record of customers thus building relationship with them; purchase management to help manage the purchase of supplies; sales management to provide a complete point of sale solution integrated to help deliver sales efficiently; and report generation.

3.3 Cloud Platform Providers

Quite a number of platform providers exist depending on the preference and choice of retail organizations. They generally provide language runtimes and data storage services for their platform. Notable among these cloud platform providers are Microsoft – with their Windows Azure cloud platform [14] [15] and Google – with its Google App Engine platform [16] [17]. With Windows Azure, applications can be built using any language, framework or tool particularly Python, Java, Ruby, Node.js and PHP. SQL Azure is a service for managing Microsoft SQL databases. The Google App Engine platform provides language runtimes for Python, Java, PHP and Go programming languages. Google Cloud SQL is the service provided for maintaining and administering MySQL databases.

3.4 Cloud Infrastructure

Cloud platform providers leverage their platforms with their own cloud infrastructure. So Microsoft and Google as well as other platform providers support their respective platforms through their cloud infrastructure, which includes: networks, servers and storage.

4 Implementation of the Proposed System

We implemented a cloud-based retail management solution called SkyRetail based in the design in Section 3. The cloud platform provider used was Microsoft with its Windows Azure platform. In order to realize the retail management system modules, we leveraged on Open Source Point of Sale – an open source POS solution developed in PHP and hosted on SourceForge repository and the Web address is given as - http://sourceforge.net/projects/opensourcepos/. The modules of the retail management system are discussed in the sub-sections that follow:

4.1 Home Page

This is the first page the user is redirected to after logging into the system. It displays a list of available modules to that particular user. It is shown in Fig. 2.

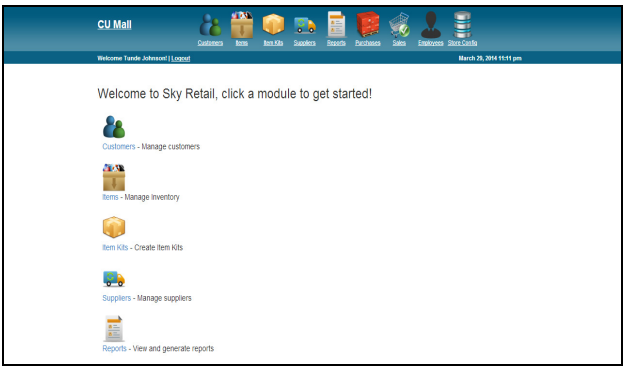


Fig. 2. Sky Retail Home Page

4.2 Customer Management Module

This administers the customer relationship management of the system. Here, the user can add new customers; view customers' details and also make changes. The Customers link on the application homepage or menu bar launches this module. The page lists out existing customers associated with the retail store. The tick box on the first column is to select customer(s) so that operations such as deleting and emailing can be carried out. Customer information can be updated by clicking the edit link on the last column. A new customer is added by filling the form displayed after pressing "New Customer" button. The search box allows the operator to search for a customer by typing a string, which could be part of the customer's name. Multiple customer information can be added to the system by clicking excel import button. The module's interface is shown in Fig. 3.

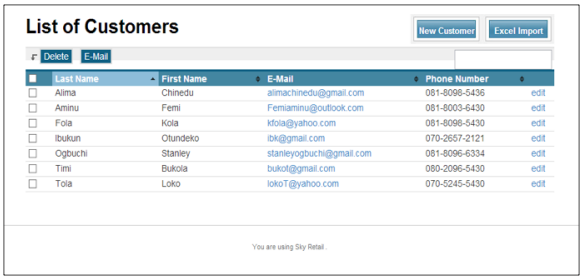


Fig. 3. Overview of the Customer Management Module

4.3 Inventory Management Module

This controls the inventory of the retail store. It provides options to add items, view the inventory details of a particular item and also make updates or changes to items. The user can group similar items into item kits. The module is depicted in Fig. 4.

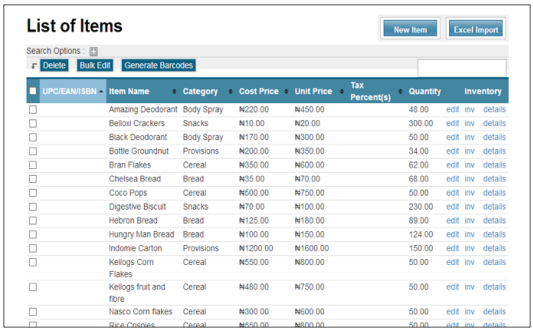


Fig. 4. Overview of the Inventory Management Module

4.4 Supplier Management Module

This provides an interface for managing the retail vendors associated with the store. The Suppliers hyperlink on the application home page or menu bar directs the user to this module. A list of existing store suppliers is displayed. Provided are different links and buttons to conduct varying operations. The tick box on the first column is used to select customer(s) for performing operations such as deleting and emailing. A supplier detail can be updated by clicking the Edit link on the last column. A new supplier is added by filling the form displayed after pressing "New Supplier" button in Fig. 5.

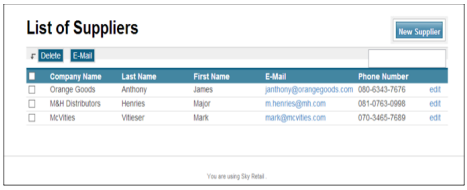


Fig. 5. Overview of the Supplier Management Module

4.5 Reports

This module makes it possible to generate various reports based on a retail store's transactions. The interface is depicted in Fig. 6. Sample reports generated by the system are shown in Figs. 7 – 8

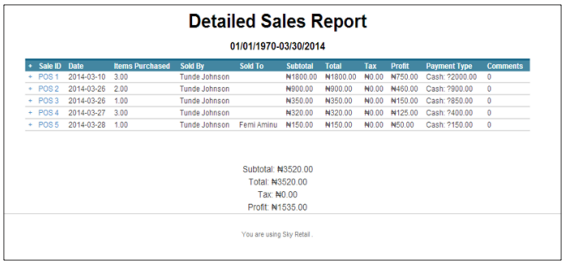


Fig. 6. This figure shows the view of detailed sales report

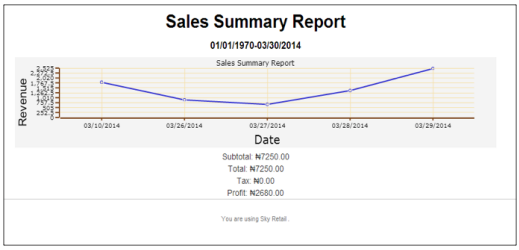


Fig. 7. This figure shows the graphical view of sales summary report

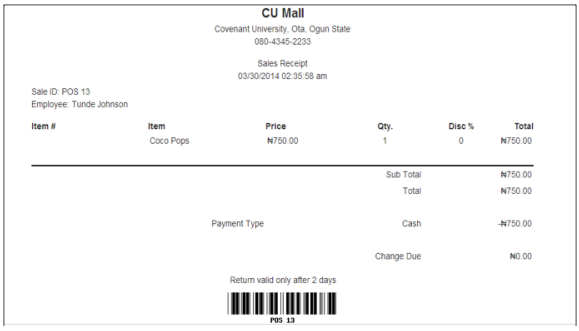


Fig. 8. This figure shows the view of Sales receipt

4.6 Purchases Module

This module manages the purchase of items from the suppliers. Clicking the "Purchase" link on the application menu bar or home page directs the user to this module. Items are added by simply searching for the item in the Find/Scan Item search box and selecting from the search dropdown. More items can be added by repeating this procedure. The sales detail for each item can be added in the text boxes provided. The supplier involved can be optionally added if the supplier exists in the system by

typing the supplier's name into Select Supplier text box. The user then selects the payment type and inputs the payment amount tendered. The Finish button is clicked to complete the purchase order. The purchase is completed and the receipt is displayed. The purchase module is depicted in Fig. 9.

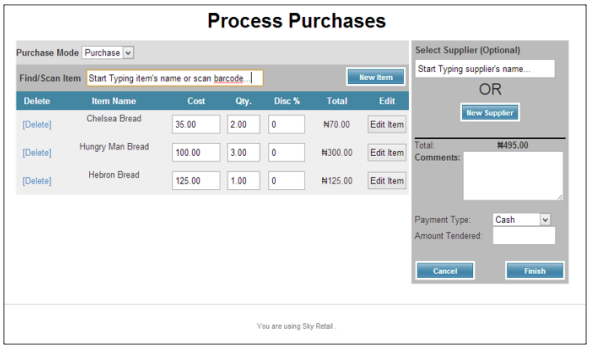


Fig. 9. This figure shows the overview of purchase module

4.7 Sales Management Module

This modules handles item sale processes. The Sales hyperlink on the application menu bar/home page launches the module. Items are added by simply searching for the item in the Find/Scan Item search box and selecting from the search drop down. More items can be added by repeating this procedure. The sales detail for each item can be added in the text boxes provided. The user can optionally choose to add the customer involved if the customer exists in the system. The user then selects the payment type and input the payment amount tendered. The Add Payment button is then clicked to submit payment. The user can decide to either complete the sale by clicking the Complete Sale button or suspend the sale by clicking the Suspend Sale button. If the sale is completed then the receipt will be displayed. Fig. 10 shows this module.

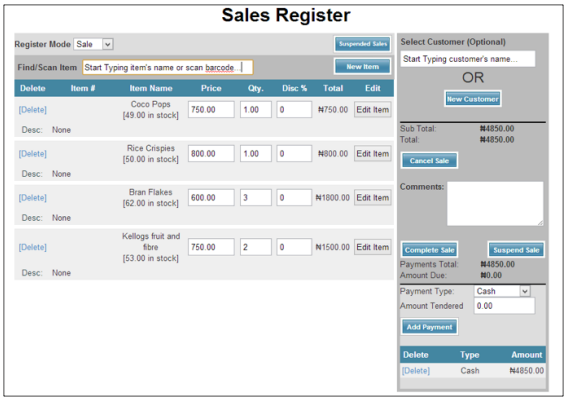


Fig. 10. This figure shows the overview of the sales module

4.8 Returns Management Module

This manages the return of the sold items of the retail store. For purchase returns, the Purchase link is clicked, then from the resulting page the user changes the mode to Returns by selecting returns from the purchase select dropdown. All necessary fields are filled and the Finish button is clicked to complete the purchase return. Clicking the Sales link and changing the register mode to returns can manage sales returns. The necessary fields are to be completed before the sales returns can be successful. This is depicted in Figs. 11 and 12.

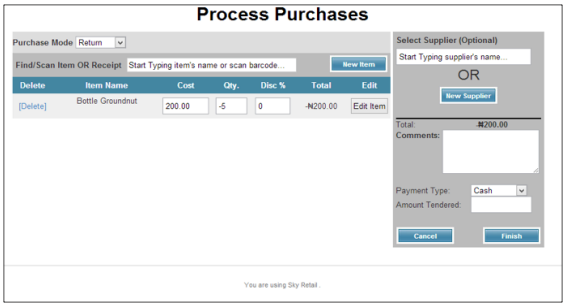


Fig. 11. This figure shows the Purchase Return module

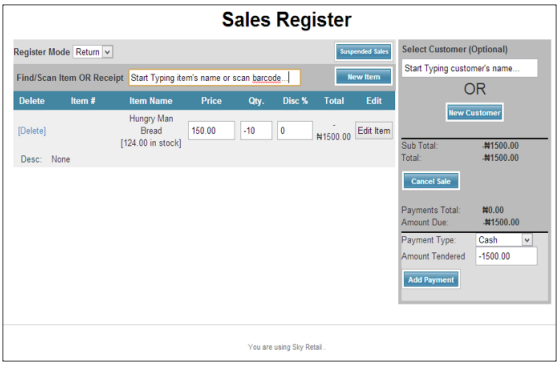


Fig. 12. This figure shows the Sales Return module

5 Conclusion

The cloud approach to implementing a retail management system adopted in this project offers massive benefits to retail organizations in managing retail operations. It will help them to overcome the high costs of running and maintaining technology infrastructure in-house. Also, the retailer can easily scale up or scale down compute resources effectively as the volume of the demand may be. Cloud computing will

allow the retail organization to focus more attention on core activities of efficiency of workflow and business operations rather that maintaining servers or systems. Furthermore, Sky retail has demonstrated the viability of the framework on which it is developed.

The retail management solution developed in this paper needs to be tested and validated by would-be users and so usability studies of the system will be conducted as part of future work. Also, as a way of improving the core features of Sky Retail we will be integrating third-party tools such as a payment gateway to process card payments and also SMS services to deliver instant messages or reports to the users.

References

1. Bajaj, C., Srivastava, N.V., Tuli, R.: Retail management, vol. 8. Oxford University Press (2005)
2. Stevenson, W.J., Hojati, M.: Operations management, vol. McGraw-Hill/Irwin, Boston (2007)
3. Andersen, R.: The rhetoric of enterprise content management (ECM): Confronting the assumptions driving ECM adoption and transforming technical communication. Technical Communication Quarterly 17, 61–87 (2007)
4. Power, D.: Supply chain management integration and implementation: a literature review. Supply Chain Management 10, 252–263 (2005)
5. Carr, N.G.: IT doesn't matter. Educause Review 38, 24–38 (2003)
6. Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., Ghalsasi, A.: Cloud computing—The business perspective. Decision Support Systems 51, 176–189 (2011)
7. Rosenthal, A., Mork, P., Li, M.H., Stanford, J., Koester, D., Reynolds, P.: Cloud computing: a new business paradigm for biomedical information sharing. Journal of Biomedical Informatics 43, 342–353 (2010)
8. Sotomayor, B., Montero, R.S., Llorente, R.M., Foster, I.: Virtual infrastructure management in private and hybrid clouds. IEEE Internet Computing 13, 14–22 (2009)
9. Moessner, K., Le Gall, F., Cousin, P.: Internet of Things Strategic Research and Innovation Agenda. Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems 7 (2013)
10. Zhang, Z., Lee, M.K., Huang, P., Zhang, L., Huang, X.: A framework of ERP systems implementation success in China: An empirical study. International Journal of Production Economics 98, 56–80 (2005)
11. Microsoft Dynamics, R.M.S., Microsoft Dynamics, R.M.S.: Overview (Date accessed: July 29th 2014) http://www.microsoft.com/en-us/dynamics/rms-overview.aspx
12. LightSpeed, Light Speed Retail (Date accessed: July 29th 2014) http://www.lightspeedretail.com
13. Epicor, Epicor Cloud Retail Software (Date accessed: July 29th 2014) http://www.epicor.com/Cloud/Pages/Cloud-Retail.aspx
14. Hill, Z., Li, J., Mao, M., Ruiz-Alvarez, A., Humphrey, M.: Early observations on the performance of Windows Azure. In: 19th ACM International Symposium on High Performance Distributed Computing, pp. 367–376 (2010)
15. Windows Azure, Microsft Cloud Platform (Date accessed: July 29th 2014), https://azure.microsoft.com/en-us/

16. Buyya, R., Yeo, C., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems 25, 599–616 (2009)
17. Google App Engine, Google App Engine: Platform as a Service (Date accessed: July 29th 2014), https://developers.google.com/appengine/

Towards a Cloud-Based Data Storage Medium for E-learning Systems in Developing Countries

Temitope Olokunde and Sanjay Misra

Covenant University, Department of Computer and Information Science, Ota, Nigeria
{temitope.olokunde,sanjay.misra}@covenantuniversity.edu.ng

Abstract. The focus of this study is to propose a cloud computing data storage model for E-Learning systems in developing countries. Cloud computing is an information technology platform that refers to services which provide data storage, collaboration and software execution hosting services via the Internet. Cloud computing is a technology trend that has a significant impact on the teaching and learning environment. The idea behind this research work therefore is to enhance the storage capacity and utilize resources of E-learning system for universities in developing countries. This platform incorporates cloud data storage medium to accommodate/store all educational content/files on the E-learning system.

Keywords: E-Learning, Cloud Computing, Data storage, University, Education, developing countries.

1 Introduction

Technology has transformed the mode of learning and created a new dimension in the educational sector. The E-Learning idea has created a connection and openness for the world irrespective of age, place, time and socio-economic nature by exploiting the web as a communication channel particularly for people in the field of education to connect distance learners with their learning resources. E-learning is one of the state-of the-art educational technologies to facilitate the learning and teaching process [1].

E-learning is about information, communication, education and learning in an online environment [2]. This aids learners in having knowledge/information at their fingertips. E-learning plays a vital role in creating a personalized learning environment that delivers learning contents to learners who have different backgrounds, interests, and locations away from a classroom in order to maximize the effectiveness of learning [3]. In a university E-learning context, students can have access to course content, video tutorials, educational e-books, online-exams, podcasts, workshop/conference videos etc. Also lecturers/instructors are able to upload course content, video tutorials and other educational materials for the student to use [4]. In maintaining an E-Learning system's data/information effectively, appropriate data storage technology must be leveraged to enable an efficient E-learning system where authorized students/lecturer can have access to information anytime and anywhere [5]. The

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 457–468, 2015.
DOI: 10.1007/978-3-662-46742-8_42

E-learning system makes learning more student-centered than instructor-centered, as learners can organize and arrange the learning contents based on their needs and suitability [6].

Cloud computing based e-learning provides continuous (anytime, anywhere and any device learning) and collaborative learning. E-learning cloud computing services is Internet-based; all of the application data must be transported based on a remote network, and data traffic surging with a very high bandwidth requirement [7]. The widely accepted benefits of cloud-based systems include: cost, device and location independence, multi-tenancy, reliability and scalability, security and environmental sustainability [8].

The use of cloud computing on universities has many benefits such as reducing universities IT complexity and cost, accessing the file storages, databases, educational resources, research applications and tools anywhere, anytime on demand [9]. The cloud-based environment supports the creation of new generation of e-learning systems, able to run on a wide range of hardware devices, while storing data inside the cloud [10].

Cloud computing is considered an internet technology for educational institutions to achieve diverse benefits that majors around resource management and effective database management [11].

The paper is structured in as follows. In the next section background of Cloud computing is summarized. Review of university e-learning in developing countries is in section 3. The proposed model and its initial implementation is in section 4 and finally conclusion drawn is in section 5.

2 Background of Cloud Computing

There are literally dozens or perhaps hundreds of definitions of cloud services that attempt to capture the essence, one that seems to have gained traction is the definition from the National Institute of Standards and Technology (NIST) which goes thus:

" Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources(e.g. networks, servers, storage applications and services) that can be rapidly positioned and released with minimal management effort or service provider interaction" [12].

Cloud computing is a new paradigm in information technology that provides an appropriate pool of computing resources with its dynamic scalability and usage of virtualized resources as a service through the Internet. With the tremendous advantages of cloud computing, we expect this technology to revolutionize the field of e-learning education. Cloud computing applications provide flexibility for all educational universities and institutions.

2.1 Cloud Types

The U.S National Institute of Standards and Technology (NIST) has a set of working definitions that separate cloud computing into two distinct set of models [12]:

-Deployment models: As shown in Figure 1 the Visual model of cloud computing above [12], the deployment models refer to the location and management of cloud's infrastructure

-Service models: These models refer to the particular services that can be accessed via cloud computing platforms

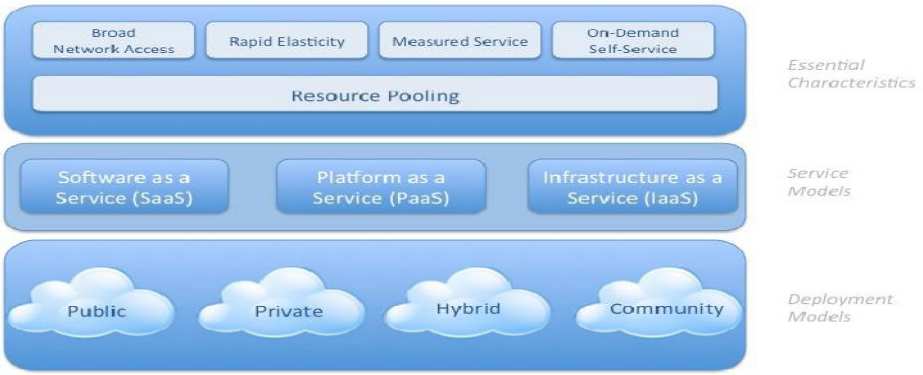


Fig. 1. Visual Model of NIST Definition of Cloud Computing. Source: (CLOUD), IEEE International Conference, pp 516-524).

2.2 Cloud Computing Deployment Models

There are four major cloud computing deployment models which include: Private cloud, Public cloud, Hybrid cloud and Community cloud. Selection of these models depends on clients' data sensitivity and management requirements [12]. An overview on these deployment models are highlighted as follows:

Public cloud: The term public does not mean that users' data is publicly visible; the infrastructure is offered via web applications as well as web services over the internet to the public or a large industry group and is owned by an organization selling cloud services. Public cloud providers such as Google or Amazon offer an access control to their clients [12].

Private cloud: the infrastructure is dedicated to a single particular organization or group. It is not shared with other organizations. Private cloud can be owned or leased. There are no additional security regulations, legal requirements or bandwidth limitations, the cloud service providers and the clients have optimized control of the infrastructure and improved security [13]. Examples of Private cloud service providers include: Amazon Virtual Private cloud, Cisco Private Cloud solutions, IBM Smart Cloud foundation and Microsoft Private cloud.

Hybrid cloud: This cloud deployment model exists due to mixed needs of n organization. It is a combination of two or more cloud service deployment models [14].

A combination of a public and private cloud is joined together for the purpose of keeping business-critical data and services in their control on a private cloud and outsourcing less-critical processing to critical processing to the public cloud. An example of a solution for creating and managing a hybrid cloud is: VMware Hybrid cloud solution.

Community cloud: Community cloud is a shared infrastructure by several organizations and supports a specific community that has shared concerns. It may be managed by the organizations or a third party and may exist at on-premises or off-premises [7].

2.3 Cloud Computing Service Models

Software as a Service (SaaS): This is a complete operating environment with applications, management and the user interface. Users can access software application hosted by the cloud vendor on pay-per use basis. Examples include Google Apps, SQL Azure.

Platform as a Service (PaaS): This is an application development tool that enables service consumers to build cloud applications that run on the hosted platform [15]. Platform is run on the cloud, at this level, the provider delivers whatever is required for both software and runtime environment. Examples include: GoGrid Cloud Center, Microsoft Azure [16].

Infrastructure as a Service (IaaS): This involves offering hardware related services using the principles of cloud computing. It is the delivery of the hardware resources such as servers, storage, data center space, networking as a service with the help of virtualization. Leading versions that provide IaaS are Amazon EC2, Amazon S3, Rackspace cloud servers and flexi scale.

2.4 Basic Concept of Data Storage in Cloud Computing

Cloud Computing enables a storage infrastructure that allows resources to be efficiently utilized, and shared. Data storage in cloud computing has the following qualities; elasticity, scalability, reliability, ease of management, ease of data access. Cloud storage system architecture consists mainly of storage layer, basic management layer, application interface layer and access layer [17]. The cloud storage allows data to be stored on multiple third party servers, rather than on the dedicated sever used in traditional networked data storage [18]. Amazon web service (AWS) is an example of a cloud service provider that offers multiple cloud-based storage options. Amazon Web Service (AWS) cloud computing platform offers some primary data storage options such as: Amazon EC2 Elastic Block Storage (EBS) volumes, Amazon Simple Storage Service (Amazon S3), Amazon EC2 Relational Databases, Amazon Relational Database Service (RDS) [19]. Web services provide huge storage capacity and scalable computing resources to end users [20].

3 Review of University E-learning Systems in Developing Countries

The adoption of Cloud based E-learning systems is yet to get a wide university audience in developing countries. Most universities in developing countries use popular E-learning platforms such as Moodle, Web-based distance learning sites. Some of these universities cut across countries like:

1. Nigeria: Some universities in Nigeria that use E-Learning platform for the academic/learning process include:
    - Covenant University: A private university that uses Moodle for academic activities like: delivering lecture notes, continuous assessment tests and assignments to her students (moodle2.covenantuniversity.edu.)
    - National Open University Nigeria (NOUN): A federal university that operates as an Open and Distance learning institution. The University provides distance learning for people through the use of internet and interactive e-learning systems.
2. South Africa: Some universities in South Africa that use E-Learning platform for the academic/learning process include:
    - University of Zululand: A public university that uses learning management system to provide the Education faculty access to e-learning. (elearn.uzulu.ac.za/moodle/edu/)
    - Cape Peninsula University of Technology: the only public university of technology in the Western Cape. This university has a center for e-learning that delivers a strategic support service to lecturers. The Learning Management system provides a secure and stable environment for teaching and learning activities. Students can receive assistance from lecturers or tutors and submit online assessments and assignments.(www.cput.ac.za/services/elearning)
3. Malaysia: Some universities in Malaysia that use E-Learning platform for the academic/learning process include:
    - University of Technology Malaysia: A public engineering and technological university that uses E-learning with features such as: news forum, school announcements, course management and embedding multimedia files.(elearning.utm.my)

The use of E-learning systems in these Universities enhances access to educational materials for learners/students, as well as an open communication channel for instructors/learners [21]. However, it is also important to have an adequate and effective storage/backup of all educational files/materials in the University; many forms of educational/academic materials need a large repository and the cloud environment can be used as a storage platform to store mass files. Cloud computing is a new paradigm of innovation in the world of Information technology, that can be used through the

internet. It is for this reason, the Cloud based E-learning system architecture is proposed. It would be of benefit for the E-learning system to be hosted on a Cloud computing technology to bridge the gap of cost on ICT infrastructure owned by any educational institutions Some researchers have proposed models that relate to cloud computing in education for universities in developing countries; these works are related to this research paper and they include:

1. NIGERIA:
   NIGEDU CLOUD: Nigedu Cloud is a model for a government-funded and administered cloud platform for e-education. Third party developers can create APIs for institution-specific customization of education-related free and open source software. The Nigedu cloud defines three roles: Cloud Service Consumer, Cloud Service Provider and Cloud Service Creator. The government is the Cloud Service Provider; provides and maintains cloud services to the consumers. The cloud consumers are tertiary institutions which includes universities, polytechnics and colleges of education in Nigeria. The cloud customers request, consume and manage cloud services some of which may include course management systems, enrollment management systems, e-examination systems etc [22].
2. INDONESIA:
   CLOUD-BASED OPEN LEARNING ARCHITECTURE: The service that runs in Open Learning portal is Student-Centered E-learning Environment-Personalization Dynamic E-learning. Learning style of the learner determines the level of learning material suitable with the learner preferences. The researchers of this work stressed the cost issue of implementing the E-learning system. They proposed a solution for these problems by adopting cloud technology and the concept of open educational resources. The result of their evaluation shows that by implementing the cloud based open learning portal, a decrease in cost is achieved [23].
3. INDIA:
   CLOUD BASED IT ARCHITECTURE FOR HIGHER EDUCATION INSTITUTE: The IaaS (Infrastructure as a Service) is the foundation of all cloud services, with PaaS(Platform as a Service) built upon IaaS and followed by SaaS(Software as a Service). IaaS contains the infrastructure resource and the hardware platforms. PaaS provides additional layer containing programming environment, middleware capabilities, database, messaging and queuing. This allows cloud users/developers to build their applications on the cloud platform. The SaaS is built upon the underlying IaaS and PaaS stacks. SaaS provides the application level capabilities to users. It includes functions for presentation, application, and provides management capabilities. [24]

4. ETHIOPIA:
   CLOUD COMPUTING FRAMEWORK FOR ETHIOPIAN HIGHER EDUCATION INSTITUTIONS: Based on the identified services and selected

clouds, the researcher proposed a Cloud Computing framework Ecosystem for Ethiopian Higher Education Institutions. The framework namely Ethiopian Universities Hybrid Cloud (EUHC) contains four layers (User Interface, SaaS, PaaS, and IaaS) and three modules (User log database, system security, and service management. The framework places as its base the hybrid cloud [25].

Cloud computing is a 21st century technology trend that cuts across all sectors of the economy and would be of immense value to the educational sector worldwide. It is with this promising benefit, the idea of incorporating cloud computing into E-learning system is birthed. This helps to enhance existing and upcoming e-learning systems particularly in developing countries.

Leaping to the cloud promises cost-effectiveness and enhanced sustainability for an e-learning system. Therefore, cloud computing should not be overlooked in the Educational sector of the economy particularly Universities and Training organizations to minimize the cost of infrastructure and human resources for development and maintenance of E-learning system.

4 Proposed Architecture

In this section, interactions that that place in the Cloud-based E-learning systems are outlined; this is used to design the system architecture (Figure 2). These include:

A. CLIENT LAYER: The clients for the E-Learning system include: students, research scholars and lecturers. These are the end users of the E-learning system; they interact with the system by logging in with authorized username and password. The user interface acts as an interface between the end-users (students, faculty, research scholars and administrators) and the cloud infrastructure. The end-users communicate with the E-Learning system via the use of browser enabled devices such as (laptop, desktop, mobile phone, iPhone, iPad).

B. APPLICATION LAYER: This is the E-learning application platform that students, research scholars use in accessing academic content, uploading academic content. The educational materials/content in the E-learning system include: Presentation videos, course files (.pdf, .ppt, .doc, .exe), mock tests and exams, assignments and case studies. The clients can interact with the application layer which is the E-learning system.

C. DATA STORAGE LAYER: Infrastructure as a Service (IaaS) involves offering hardware related services using the principles of cloud computing. These clouds include storage devices (database or disk storage) or virtual servers. This layer maintains and manages the resources of the virtualized university infrastructure by means of the components such as: live migration. The IT management team is able to control and manage the E-learning systems in terms of the operating systems, applications, storage and network connectivity. The public cloud deployment model: Amazon web service is used in storing data for the E-learning system. The IaaS therefore serves as a host for the E-learning application which helps in effective resource utilization.

4.1 UML Diagrams Depicting System

Unified Modeling Language (UML) is an international industry standard graphical notation for describing software analysis and designs. The UML diagrams used in depicting the procedures involved in the Cloud-based E-learning system includes: Use Case, Sequence and Class diagram, and they are depicted as follows:

Use Case Diagram (Fig.3) shows the typical interaction that takes place in a Cloud-based E-Learning System:

Actor 1(Administrator): This is the person in charge of the Cloud-based E-learning system who authorizes lecturers and students

Actor2 (Lecturer): Here, the lecturer is able to upload educational content for his/her students in related departments within the university. The lecturer can also assess assignments and give mock tests/exams to the students for easy access towards the learning process

Actor3(Student): The students on the University's E-learning system are able to communicate with lecturers, download presentation videos, and download files such as .pdf, .doc, .ppt, .exe etc and assignments.

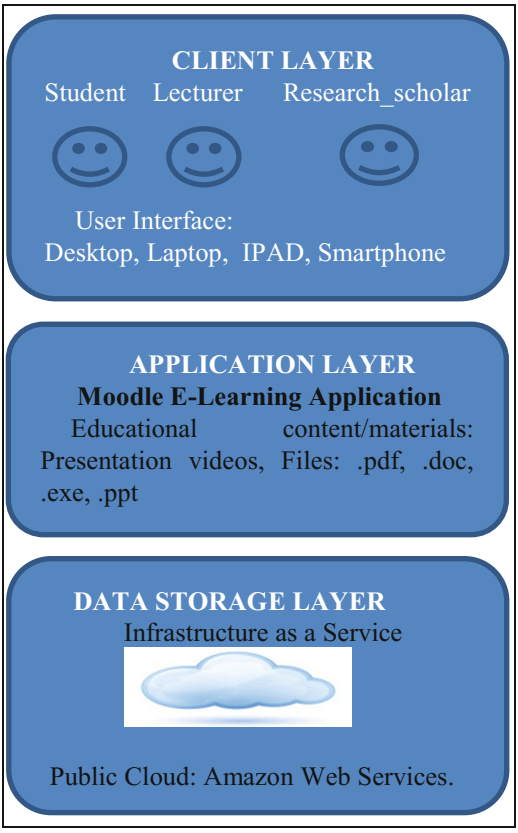


Fig. 2. Diagram depicting Cloud-based E-learning System

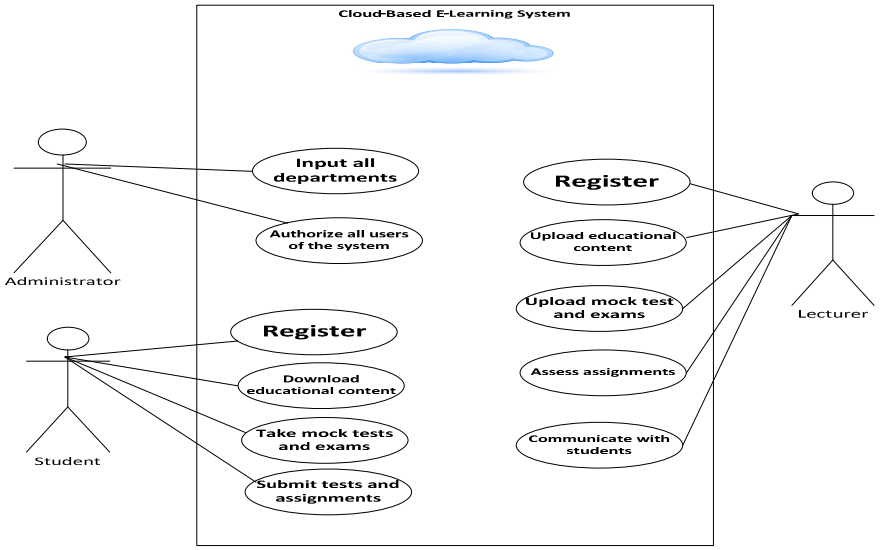


Fig. 3. Use Case Diagram

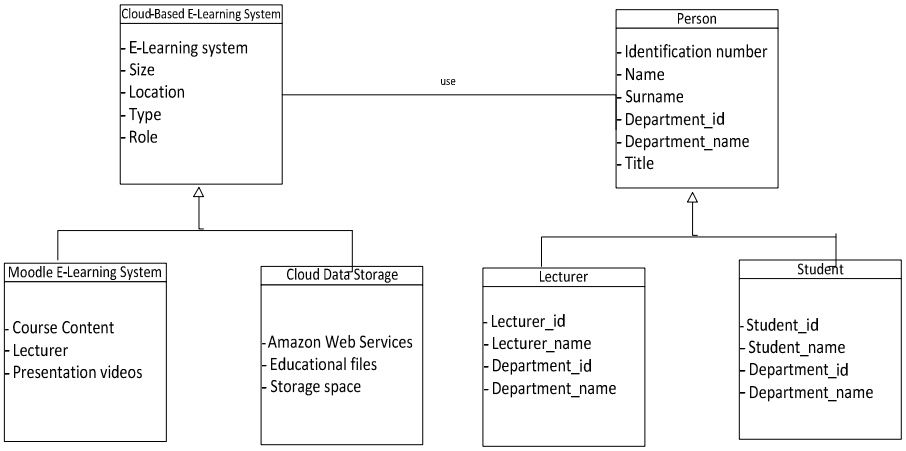


Fig. 4. Class Diagram

4.2 Class Diagram

The class diagram in Figure 4 shows the relationship between Cloud-Based E-learning systems and Persons. The Cloud Based E-learning system should possess attributes such as: identification number, name, surname, department name, department_id, and title of the person. This attributes enable smooth transition of communication and interaction within the E-learning system in the University. The E-learning system contains: course content, lecturer, presentation videos which are the major educational materials within the academic environment. The cloud storage medium used in this paper is on

the services provided by Amazon web services; a service that provides cloud storage such as: Amazon Relational Database Service, Amazon S3; a redundant and fast cloud storage service that provides public access to files.

4.3 Implementation

The cloud computing service model being used in this system is the Infrastructure as a Service (IaaS) model because it focuses on the data storage platform for the University E-Learning system. The following shows the implementation architecture for implementing and deploying a cloud-based data storage medium for E-learning systems: PUBLIC MODEL; AMAZON WEB SERVICES (AWS): Amazon Web Services provides a highly scalable and reliable cloud hosting environment for an E-learning system. Amazon RDS is a managed Relational Database Service that automatically patches the database software and backs up database and supports point-in-time recovery [14]. The Amazon Web Service (AWS) Cloud which is being considered in this paper, offers many ways to support data storage [26].

Adopting cloud computing in e-learning system helps minimize cost through a pay-as-you-use model, allows users with minimal hardware requirements to access e-learning services [26].

4.4 Discussion

Over the years, E-learning has proven to be a technology platform that facilitates the learning process for learners, as learners can read learning materials on their own without much supervision via the use of internet-enabled mobile phones, laptops, iPads, Personal Digital Assistance or collecting materials from other people using Compact discs, USB storage device etc. The use of Moodle enables students to download lecture notes, presentation videos, submit assignments, communicate with lecturers, join discussion forums, read announcements at their pace [11]. Although cloud models have been proposed by other researches, we show a comparison with others and draw out a model. The E-learning system would therefore be hosted on a public cloud provided by Amazon Web Services which aids resource management, backup, recovery and effective data storage. The cloud platform enables a repository that can store large educational content/data so Universities do not have to worry about losing data.

5 Conclusion

E-Learning Systems plays a vital role in Universities for students and lecturers/instructors, as learning can take place anytime and anywhere. Large-scale applications are data-intensive and effective and efficient storage mechanisms should be put in place. The use of a Cloud is gradually becoming a popular on-demand framework for supporting and storing intensive data. It would be of immense benefit to incorporate Cloud computing as a data storage platform for learning/educational resources across Universities in developing countries.

References

1. Nguyen, T.D., Nguyen, T.M., Pham, Q.-T., Misra, S.: Acceptance and Use of E-Learning Based on Cloud Computing: The Role of Consumer Innovativeness. In: Murgante, B., Misra, S., Rocha, A.M.A.C., Torre, C., Rocha, J.G., Falcão, M.I., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) ICCSA 2014, Part V. LNCS, vol. 8583, pp. 159–174. Springer, Heidelberg (2014)
2. Hammad, R., Odeh, M., Khan, Z.: Towards a Generic Requirements Model for Hybrid and Cloud-based E-learning Systems. In: IEEE International Conference on Cloud Computing Technology and Science, pp. 106–111. Bristol (2013)
3. Phankokkruad, M.: Implementation of Cloud Computing. In: International Conference on Computer and Information Science, pp. 7–11. Kuala, Lempeu (2012)
4. Holbl, M., Welzer, T.: Students feedback and communication habits using moodle. Electronics & Electrical Engineering, Computers & Applied Sciences 6, 63–66 (2010)
5. Jin, S.: Da hongYing, N.: Design of an online learning platform with moodle. In: Computer Science & Education (ICCSE) 7th International Conference, Melbourne, Australia, pp. 1710–1714 (2012)
6. Fasibuddin, H., Skinner, G., Athauda, R.: A Holistic Review of Cloud-based E-learning System. In: IEEE International Conference on Teaching Assessment and Learning for Engineering (TALE), Hong Kong, pp. 6–11 (2012)
7. Anwar, M., Masud, H., Huang, X.: A Novel Approach for Adopting Cloud-based E-learning System. In: IEEE/ACIS 11th International Conference on Computer and Information Science, China, pp. 37–42 (2012)
8. Zahraa, A.M., Ayman, F.: Cloud Computing Educational Environment. In: European Mediterranean & Middle Eastern Conference on Information Systems, pp. 1–6. Winsdor, United Kingdom (2013)
9. Madhumathi, C., Ganapathy, G.: An academic cloud framework for adapting e-Learning in universities. International Journal of Advanced Research in Computer and Communication Engineering 2, 4480–4484 (2013)
10. Sonwar, M.D., Xiaodi, H.M.: An E-learning System Architecture Based on Cloud Computing. World Academy of Science, Engineering and Technology 6, 738–742 (2012)
11. Pardeshi, V.: Architecture and Adoption Model for Cloud in Higher Education: Indian Perspective. Journal of Industrial Engineering and Management Science 4(2), 43–47 (2014)
12. Rozznyai, S., Alexander, S., Yurdaer, D.: Large-scale Distributed Storage System for Business Provenance. In: IEEE International Conference on Cloud Computing (CLOUD), Washington DC, pp. 516–524 (2011)
13. Metri, P., Sarote, G.: Privacy Issues and Challenges in Cloud Computing. International Journal of Advanced Sciences and Technologies (IJAEST), 1–6 (2011)
14. Phankokkruad, M.: Implementation of Cloud Computing. In: International Conference on Computer and Information Science, pp. 7–11. Kuala, Lempeu (2012)
15. William, Y.C., Hosame, A., Jessica, S.: Transforming Enterprise Cloud Services. Springer Dordrecht Heidelberg London New York (2010)
16. CSA, Security Guidance for Critical Areas of Focus in Cloud Computing V2.1 Cloud Security Alliance (2009), `http://www.cloudsecurityalliance.org` (accessed: July 08, 2011)
17. Kun, L., Long-jiang, D.: Research on cloud data storage technology and its implementation. In: International Workshop on Information and Electronics Engineering (IWIEE), vol. 29, pp. 133–137 (2012)

18. Arjun, K., et al.: Secure storage and access of data in cloud computing. In: Convergence, I.C.T. (ed.) ICT Convergence (ICTC) International Conference, pp. 336–339. IEEE, Jeju Island (2012)
19. Talasila, S., Pavan, K., Subrahmanyam, K.: A generalized cloud storage architecture with backup technology for any cloud storage providers. International Journal of Computer Application 2, 256–263 (2012)
20. Momin, B.F., Girish, B.: Distributing, ensuring and recovery of data stored in cloud. International Journal of Computer Science and Mobile Computing 3, 1–7 (2014)
21. Nahman, E-learning: a research note, http://www.namahn.com/resources/documents/e-learning.pdf
22. Ezenwoke, A., Omoregbe, N., Ayo, C.K., Sanjay, M.: NIGEDU CLOUD: Model of a national e-education cloud for developing countries. In: International Conference on Electronic Engineering and Computer Science, vol. 4, pp. 77–78 (2013)
23. Selviandro, N., Survani, M., Hasibaun, Z.A.: Open learning optimization based on cloud technology. In: Advanced Communication Technology (ICACT) 16th International Conference, pp. 541–546 (2014)
24. Blezu, C., Popa, E.M.: E-learning and its prospects. In: 12th WSEAS International Conference on COMPUTERS, Heraklion, Greece, pp. 297–302 (2008)
25. Saidhbi, S.: A cloud computing framework for ethiopian higher education institutions. Journal of Computer Engineering (IOSRJCE) 6, 1–9 (2012)
26. Chiu, D., Agrawal, G.: Evaluating Caching and Storage Options on the Amazon Web Services Cloud. In: 11th IEEE/ACM International Conference on Grid Computing(GRID), pp. 17–24. Brussels (2010)

Fast and Efficient Parallel Computations Using a Cluster of Workstations to Simulate Flood Flows

Sudi Mungkasi[1] and J. B. Budi Darmawan[2]

[1] Department of Mathematics, Faculty of Science and Technology,
Sanata Dharma University, Mrican, Tromol Pos 29, Yogyakarta 55002, Indonesia
sudi@usd.ac.id
[2] Department of Informatics, Faculty of Science and Technology,
Sanata Dharma University, Mrican, Tromol Pos 29, Yogyakarta 55002, Indonesia
b.darmawan@usd.ac.id

Abstract. A strategy is proposed for fast and efficient parallel computations using a cluster of workstations to simulate flood flows. We implement ANUGA software to conduct flood simulations. ANUGA solves the two-dimensional shallow water equations using a finite volume method. ANUGA can be run in either sequential or parallel. We focus our work on parallel computations. Our computation scenarios are implemented on a cluster of workstations. Parallel computation results are assessed based on execution time and efficiency. The strategy that we propose gives a significant advantage when we want to get fast and efficient computations using a cluster of workstations.

Keywords: ANUGA software, flood flow, parallel programming.

1 Introduction

Flood flow simulations are needed for an early warning system before the real event. The system warns the community when there in an indication of an upcoming flood. To decide whether a warning should be issued or not, simulations of floods should be conducted beforehand.

Fast and efficient flood simulations are always desired in addition to accurate results [1–3]. In terms of computation, we need to balance between speed, efficiency, and accuracy of simulations. Accuracy of simulations has been discussed thoroughly in a number of references, such as Guinot [5], LeVeque [8] and Mungkasi [11]. In contrast, speed and efficiency of flood flow simulations still need to be enhanced. This paper limits the discussion to speed and efficiency of flood flow simulations using ANUGA software.

To gain a high speed computation, running a simulation sequentially is not the best option. Therefore, we run simulations in parallel. Speed and efficiency are considered and analysed in order to assess parallel computation results. This paper gives a strategy to get fast and efficient parallel computations using a cluster of workstations to simulate flood flows.

The rest of this paper is organised as follows. We recall the mathematical model governing water flows in Section 2. The model is solved numerically in Section 3.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 469–477, 2015.
DOI: 10.1007/978-3-662-46742-8_43

The numerical method is implemented in parallel into ANUGA software, as described in Section 4. We provide computational results and discussion about the results in Section 5. Finally, we draw some concluding remarks in Section 6.

2 Governing Equations

In this section we recall the mathematical model governing shallow water flows. We refer to Mungkasi and Roberts [10] for the description of the model. More detailed explanation of the shallow water equations can be found in other work, such as [5, 15, 17].

The vector form of the two-dimensional shallow water equations is

$$\mathbf{q}_t + \mathbf{f}(\mathbf{q})_x + \mathbf{g}(\mathbf{q})_y = \mathbf{s}\,. \tag{1}$$

Here $\mathbf{q} = [h\ uh\ vh]^T$ is the vector consisting of water depth h, x-momentum uh, and y-momentum vh. Variables u and v represent velocities in the x- and y-direction respectively. Functions $\mathbf{f}(\mathbf{q})$ and $\mathbf{g}(\mathbf{q})$ are fluxes in the x- and y-direction respectively, where

$$\mathbf{f}(\mathbf{q}) = \begin{bmatrix} uh \\ u^2h + \frac{1}{2}gh^2 \\ uvh \end{bmatrix} \quad \text{and} \quad \mathbf{g}(\mathbf{q}) = \begin{bmatrix} vh \\ vuh \\ v^2h + \frac{1}{2}gh^2 \end{bmatrix}. \tag{2}$$

The right hand side $\mathbf{s}$ of (1) is the source term given by

$$\mathbf{s} = \begin{bmatrix} 0 \\ -gh\,(z_x + S_{fx}) \\ -gh\,(z_y + S_{fy}) \end{bmatrix} \tag{3}$$

where $z(x, y)$ is the bed topography, and $S_f = \sqrt{S_{fx}^2 + S_{fy}^2}$ is the bed friction following Manning's resistance law

$$S_{fx} = \frac{u\eta^2\sqrt{u^2+v^2}}{h^{4/3}} \quad \text{and} \quad S_{fy} = \frac{v\eta^2\sqrt{u^2+v^2}}{h^{4/3}}. \tag{4}$$

The constant η is the Manning resistance coefficient. The notation g represents the acceleration due to gravity.

We integrate (1) over an arbitrary closed and connected spatial domain Ω having boundary Γ and apply the Gauss divergence theorem to the flux terms to get the integral form

$$\frac{\partial}{\partial t}\int_\Omega \mathbf{q}\,d\Omega + \oint_\Gamma \mathbf{F}\cdot\mathbf{n}\,d\Gamma = \int_\Omega \mathbf{s}\,d\Omega\,. \tag{5}$$

Here $\mathbf{F} = [\mathbf{f}(\mathbf{q})\ \mathbf{g}(\mathbf{q})]^T$ is the flux vector, $\mathbf{n} = [\cos(\theta)\ \sin(\theta)]^T$ is the outward normal vector of the boundary, and θ is the angle between $\mathbf{n}$ and the x-direction.

The shallow water wave equations admit the rotational invariance property, which implies [17]

$$\mathbf{F} \cdot \mathbf{n} = \mathbf{T}^{-1}\mathbf{f}(\mathbf{T}\mathbf{q}) \tag{6}$$

where $\mathbf{T}$ is the transformation matrix

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{bmatrix} . \tag{7}$$

Therefore, equation (5) can now be written as

$$\frac{\partial}{\partial t}\int_{\Omega} \mathbf{q}\, d\Omega + \oint_{\Gamma} \mathbf{T}^{-1}\mathbf{f}(\mathbf{T}\mathbf{q})\, d\Gamma = \int_{\Omega} \mathbf{s}\, d\Omega\,. \tag{8}$$

This integral form (8) of the shallow water equations helps in the construction of a numerical method that we discuss in the next section.

3 Numerical Method

We solve the shallow water equations numerically using a finite volume method. The spatial domain is discretised into a finite number of triangular cells.

The semi-discrete scheme of the finite volume method corresponding to equation (8) is

$$\frac{d\mathbf{q}_i}{dt} + \frac{1}{A_i}\sum_{j\in\{0,1,2\}} \mathbf{H}_{ij}\, l_{ij} = \mathbf{s}_i\,. \tag{9}$$

Here $\mathbf{q}_i$ is the vector of conserved quantities averaged over the ith cell, A_i is the area of the ith triangle (cell), $\mathbf{s}_i$ is the source term associated with the ith cell, $\mathbf{H}_{ij}$ is the outward normal flux of material across the ij edge, and l_{ij} is the length of the ij edge. The ij edge is the interface between the ith and jth cells. Equation (9) is implemented on each triangular cell. The summation in sigma notation of equation (9) goes for $j \in \{0, 1, 2\}$, as each cell has three edges. Readers interested in the algorithm to compute the numerical flux $\mathbf{H}_{ij}$ should consult the work of Roberts et al. [10, 15].

The finite volume method (9) has been implemented into ANUGA software. ANUGA is a free and open source software developed by the Australian National University (ANU) and Geoscience Australia (GA). It can be downloaded from its official website `http://anuga.anu.edu.au`. It is available to be installed on Windows and Linux machines. ANUGA can be run in either sequential or parallel. It has been widely used for simulating flood inundations [6, 7, 12–14].

4 Parallel Programming

In this section we compare flood simulation results from several scenarios of parallel computations. We implement MPI Personal Computer cluster for parallel

programming with ANUGA. Parallel implementation of ANUGA is described in [19].

We are interested in cluster computations using PCs interconnected by network technology. These computations have great advantages, such as high performance for low price, system scalability, and rapid adjustment to new technological advances [16]. The availability of Gigabit Ethernet supports those advantages. Gigabit Ethernet has speed up the communication performance of PC cluster using MPI than ever before such as Fast Ethernet, Myrinet and Scalable Coherent Interface [9]. One of the most popular communication protocols for PC cluster to support parallel programming is Message Passing Interface (MPI). MPI is a library specification for message-passing in the network [20] and one of its implementations is MPICH. MPICH is a high performance and widely portable implementation of the MPI standard (see [21] for more details about MPICH).

ANUGA is developed based on Python programming language. It supports parallel programming with MPI using pypar as the interface between MPI and Python. OPENMPI and MPICH2 are supported by pypar [19]. We use MPICH2 for our parallel computations as the implementation of MPI.

Foster [4] proposed efficiency to be a metric to evaluate parallel algorithm performance. They [4] noted that execution time varied with problem size, but efficiency is independent of problem size. However execution time is also an important issue in computations. Therefore, in this paper we consider the execution time as well as the efficiency to evaluate our parallel algorithm scenarios.

The relative speedup S_{relative} is formulated as

$$S_{\text{relative}} = \frac{T_1}{T_p}, \tag{10}$$

where T_1 is the execution time on one processor and T_p is the time on P processors. The relative speedup means the factor by which execution time is reduced on P processors. Relative efficiency E_{relative} is defined as

$$E_{\text{relative}} = \frac{T_1}{PT_p}. \tag{11}$$

Note that equation (10) is equivalent to $S_{\text{relative}} = PE_{\text{relative}}$. This means that the speedup is known if and only if the efficiency is known.

5 Numerical Results

In this section we present main results of our research.

For a preliminary illustration, we present simulation results of the Cairn tsunami flood inundations. The numerical setting is available in the current ANUGA software. Representatives of simulation results are given in Figures 1 and 2. Figure 1 shows the initial condition of the Cairn coastal region in Australia before an incoming flow from the ocean exists. At time $t > 0$ we assume

that there is a fixed incoming flow from the ocean, which results in flood inundations around the coastal region. This is illustrated in Figure 2 showing the situation at time $t = 5000$ seconds simulated real time.

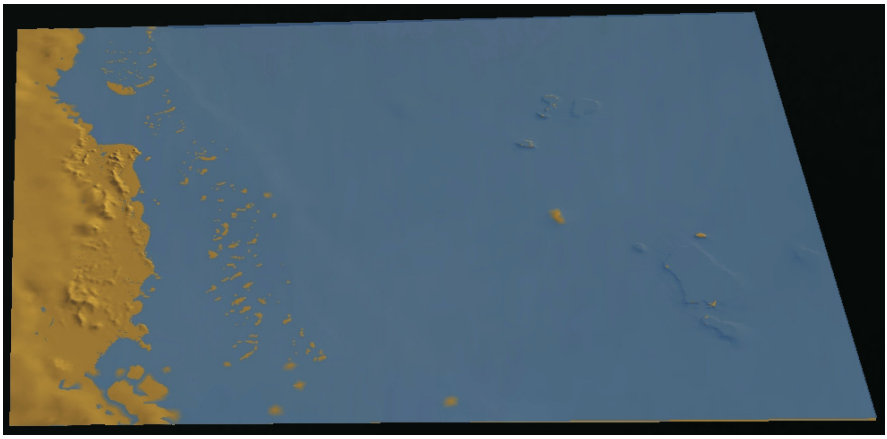


Fig. 1. Initial condition of the Cairn coastal region. Before the simulation starts, water is at rest.

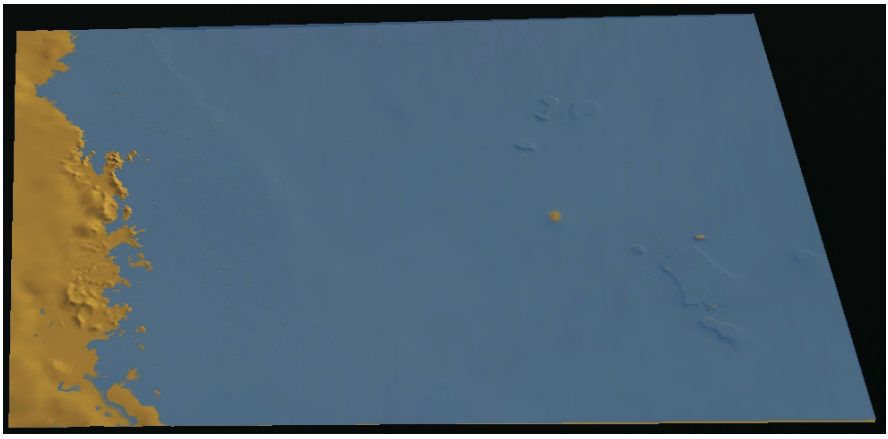


Fig. 2. Simulation result after 5000 seconds of a fixed incoming water flow from the ocean boundary. Flood inundation occurs at the Cairn coastal region.

Now we focus on our main results of parallel computations for flood flows. Simulations are conducted in a cluster of workstations in Sanata Dharma University. The cluster consists of 16 PCs with i3 processors, that is one PC has one i3 processor. Note that an i3 processor has 2 cores, and an i3 processor has 4 threads. When we simulate flood flows for our research, all 16 PCs are turned on

with Linux operating system. For the network communication we use a one Gigabit Ethernet. In addition, for the RAM, we have four gigabytes (4 GB RAM) in each PC. We refer to the work of Foster [4] and Yang et al. [18] for concepts of parallel programming.

We have conducted 20 simulations for running ANUGA software in order to simulate flood flows. One simulation is run in squential. The others are run in parallel. We consider scenarios of 1, 2, 4, 8, and 16 PCs for our computations. Using MPICH we can distribute the computation processes in parallel. For each scenario, we consider 1, 2, 3, and 4 processes on each PC.

For a flood flow problem we record the total time for each simulation in Table 1. We see that running the simulation into either 3 processes or 4 processes does not make the computation faster than just 1 or 2 processes. This is reasonable because an i3 processor has only 2 cores even though it has 4 threads. The number of cores is more crucial in computations. The total time for all simulations are viewed in Figure 3.

Table 1. Total time for a simulation scenario using 1 Gbps

Number of PCs	1 process	2 processes	3 processes	4 processes
1	2828.41	2255.50	2455.60	2505.65
2	1653.14	1311.79	1370.35	1457.07
4	1060.57	819.40	806.83	785.04
8	710.20	495.09	578.32	550.70
16	431.46	383.81	472.92	543.72

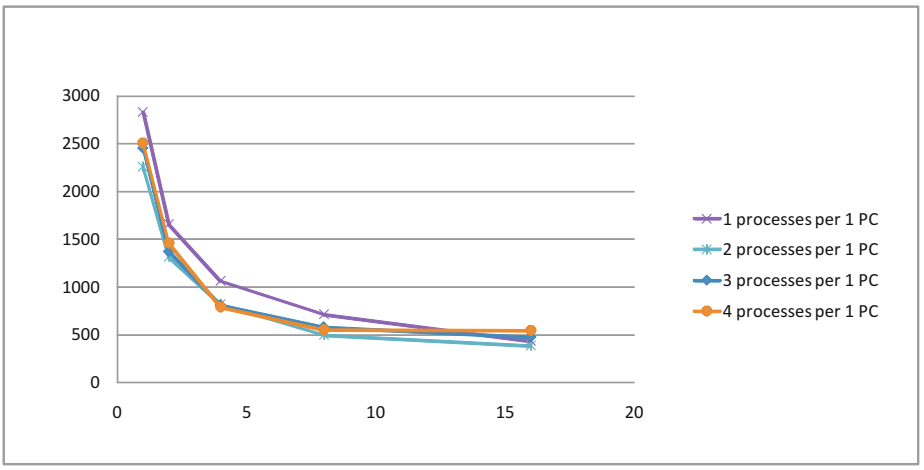


Fig. 3. Total time elapsed in a simulation scenario for several computational settings. The horizontal axis is the number of PCs. The vertical axis is the total time used for computations.

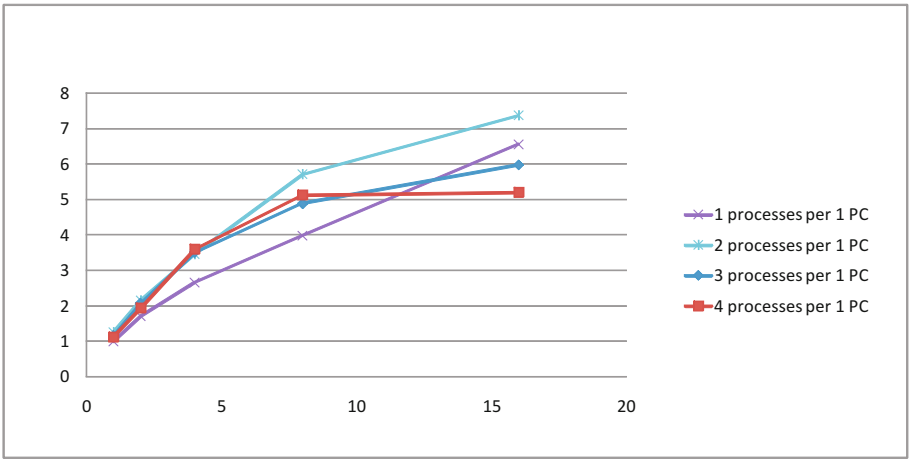


Fig. 4. Speed up obtained from parallel computations for a simulation scenario. The horizontal axis is the number of PCs. The vertical axis is the speed up.

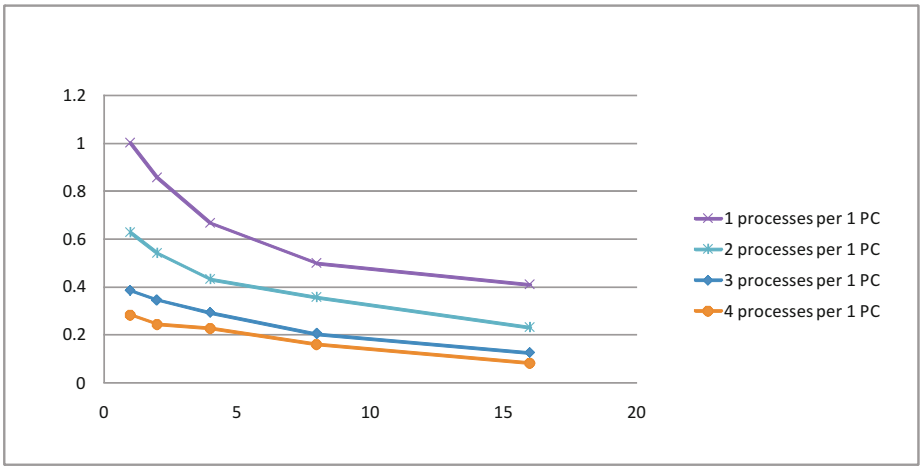


Fig. 5. Efficiency of parallel computations for a simulation scenario. The horizontal axis is the number of PCs. The vertical axis is the efficiency.

For the speed up we observe that more number of PCs leads to faster computation, which is reasonable. This is shown in Figure 4. However more number of PCs makes the computation less efficient, as shown in Figure 5. This can be understood as follows: physically the resources are wasted if the number of PCs is too many, and mathematically the divisor in equation (11) gets larger when P increases, which makes the efficiency low. From the speed and efficiency, we recommend that we use only one core in each i3 machine. Again, using all two

cores in each i3 machine results in less efficient computation than our recommendation. We believe that this is due to some background tasks or other processes. One of the background tasks is the operating system turned on for each machine (note that one of the two cores in each i3 machine has run for the operating system already). Another task is the MPI process distributing the computations in parallel.

6 Conclusion

We have simulated several scenarios of parallel computations for flood flows. We have obtained a strategy for fast and efficient parallel computations using a cluster of workstations consisting of PCs with i3 processors. The strategy is to use only one core in each i3 machine in order to get fast and efficient computations. Note that employing all cores in each i3 machine leads to faster computations but less efficient. For future direction, we will investigate some strategies for fast and efficient parallel computations using other types of clusters.

Acknowledgements. The authors thank Professor Stephen Gwyn Roberts at Mathematical Sciences Institute, The Australian National University, Canberra, Australia for stimulating discussions. The first author also thanks The Australian National University and SMERU Research Institute for the Indonesia Project research grant.

References

1. Baldock, T.E., Barnes, M.P., Guard, P.A., Hie, T., Hanslow, D., Ranasinghe, R., Gray, D., Nielsen, O.: Modelling Tsunami Inundation on Coastlines with Characteristic Form. In: 16th Australasian Fluid Mechanics Conference, pp. 939–942. AFMS (2007)
2. Cousins, W.J., Power, W.L., Destegul, U.Z., King, A.B., Trevethick, R., Blong, R., Weir, B., Miliauskas, B.: Earthquake and Tsunami Losses from Major Earthquakes Affecting the Wellington Region. In: 2009 New Zealand Society for Earthquake Engineering Conference, Paper Number 24. NZSEE (2009)
3. Davies, G.: A Well-Balanced Discretization for a Shallow Water Inundation Model. In: 2011 International Congress on Modelling and Simulation (MODSIM), pp. 2824–2830. MSSANZ (2011)
4. Foster, I.: Designing and Building Parallel Programs: Concepts and Tools for Parallel Software Engineering. Addison-Wesley, Boston, Massachusetts (1995)
5. Guinot, V.: Wave Propagation in Fluids: Models and Numerical Techniques. ISTE, London - Wiley, Hoboken (2008)
6. Jakeman, J.D., Bartzis, N., Nielsen, O., Roberts, S.: Inundation Modelling of the December 2004 Indian Ocean Tsunami. In: 2007 International Congress on Modelling and Simulation (MODSIM), pp. 1667–1673. MSSANZ (2007)
7. Jakeman, J.D., Nielsen, O.M., Putten, K.V., Mleczko, R., Burbidge, D., Horspool, N.: Towards Spatially Distributed Quantitative Assessment of Tsunami Inundation Models. Ocean Dynamics 60, 1115–1138 (2010)

8. LeVeque, R.J.: Finite Volume Methods for Hyperbolic Problems. Cambridge University Press, Cambridge (2002)
9. Mache, J.: An Assessment of Gigabit Ethernet as Cluster Interconnect. In: 1st IEEE Computer Society International Workshop on Cluster Computing. IEEE (1999)
10. Mungkasi, S., Roberts, S.G.: A Finite Volume Method for Shallow Water Flows on Triangular Computational Grids. In: 2011 International Conference on Advanced Computer Science and Information System (ICACSIS), pp. 79–84. IEEE (2011)
11. Mungkasi, S.: A Study of Well-Balanced Finite Volume Methods and Refinement Indicators for the Shallow Water Equations. PhD Thesis, Australian National University, Canberra, Australia (2012)
12. Mungkasi, S., Roberts, S.G.: Predictions on Arrival Times of Water of the St. Francis Dam Break Flood Using ANUGA. In: 2013 International Congress on Modelling and Simulation (MODSIM), pp. 304–309. MSSANZ (2013)
13. Nielsen, O., Sexton, J., Gray, D., Bartzis, N.: Modelling Answers Tsunami Questions. AusGeo News 83, 1–5. Geoscience Australia (2006)
14. Roberts, S.G., Nielsen, O.M., Jakeman, J.: Simulation of Tsunami and Flash Floods. In: Modeling, Simulation and Optimization of Complex Processes, pp. 489–498. Springer, Berlin (2008)
15. Roberts, S., Nielsen, O., Gray, D., Sexton, J.: ANUGA User Manual. Geoscience Australia and Australian National University, Canberra (2010)
16. Sterling, T.: Beowulf Cluster Computing with Linux. MIT Press, Cambridge, Massachusetts (2001)
17. Tan, W.Y.: Shallow Water Hydrodynamics. Elsevier, Amsterdam (1992)
18. Yang, J., Shen, W., Yan, G.: Construct COW Based on MPICH in Linux Environment. In: 1st International Workshop on Education Technology and Computer Science. IEEE (2009)
19. AnugaParallel:
`https://anuga.anu.edu.au/wiki/AnugaParallel`
20. Message Passing Interface (MPI) Standard:
`http://www.mcs.anl.gov/research/projects/mpi`
21. MPICH High-Performance Portable MPI:
`http://www.mpich.org`

Strategic Planning

A Simulation Model for Strategic Planning in Asset Management of Electricity Distribution Network

Erma Suryani[1], Rully Agus Hendrawan[1], Eka Adipraja Philip Faster[1], and Lily Puspa Dewi[2]

[1] Information Systems Department-Faculty of Information Technology, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia
[2] Informatics Department-Faculty of Industrial Technology-Petra Christian University, Surabaya, Indonesia
erma.suryani@gmail.com

Abstract. Asset management of electricity distribution network is required in order to improve the network reliability so as to reduce electricity energy distribution losses. Due to strategic asset management requires long-term predictions; it would require a simulation model. Simulation of asset management is an approach to predict the consequences of long-term financing on maintenance and renewal strategies in electrical energy distribution networks. In this research, the simulation method used is System Dynamics based on consideration that this method enables us to consider internal and external influenced factors. To obtain the model parameter, we utilized PLN Pamekasan for the case study. The results showed the reduction of low voltage network assets condition on average in the range 6% per year, the average decline in the transformer condition is approximately 6.6% per year, and the average decline in the condition of medium voltage network assets is approximately 4.4% per year. In general, the average technical losses average of 1,359,981.60 KWH / month or about 16,319,779.24 KWH / year.

1 Introduction

1.1 Background

Electricity energy losses are losses incurred in the transfer of electricity energy through the distribution network. These losses increase along the channel of distribution and depend on the amount of transferred power. Some of the factors that influence the distribution losses are: network configuration, utilization, load profile, as well as power factor [1]. Asset management is required in order to improve the network reliability as well as to reduce the losses of electricity distribution.

There are four strategies in asset management as seen in Fig.1 [2]. Corrective Maintenance treatment is carried out when the network is impaired, while the Time Base Maintenance is maintenance and inspections that are conducted periodically at certain times. Furthermore, Condition Based Maintenance is a condition in which the

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 481–492, 2015.
DOI: 10.1007/978-3-662-46742-8_44

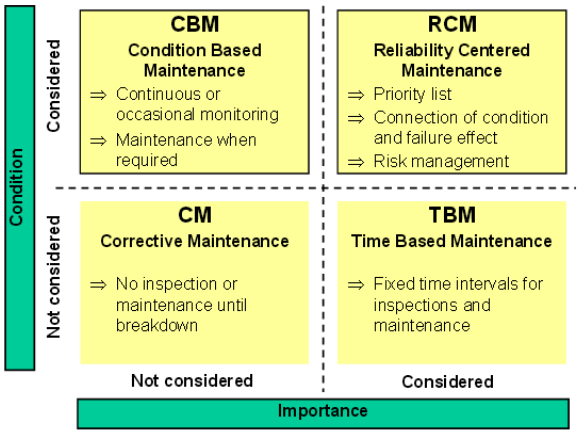


Fig. 1. Network Maintenance Strategy Classification [2]

asset is continuously monitored and conducted in accordance with the purposes of treatment. Finally, Reliability Centered Maintenance is asset maintenance that will be conducted based on priorities by considering the assets conditions and risks.

In this research, we used "Reliability Centered Maintenance" strategy which is very suitable to be applied in the area of Madura - Indonesia (Pamekasan) so that the network asset management can be performed based on priorities of the loss distribution of electrical energy which are prevalent in low voltage network (LVN). The purpose of this research is to develop a system dynamics model that can analyze the condition of the electricity distribution assets in deciding the effective and efficient asset maintenance policy.

1.2 Problem Solving Approach

In electricity energy distribution, asset management can be defined as a systematic process of the operation, maintenance, and improvement of distribution network reliability by combining the practices of "reengineering" and economic analysis [3].

In general, distribution network assets can be grouped into two parts, namely primary asset and secondary assets. The primary assets include air ducts, power transformer, high voltage, medium and low voltage cables. The secondary assets include power system protection relays, power meters, and control infrastructure. Due to strategic asset management requires long-term predictions; it would require a simulation model for asset management. Simulation asset is an approach to predict the consequences of long-term financing strategy in network maintenance and renewal of the electrical energy [4].

In this research, we utilized system dynamics framework to learn about the dynamic behavior of a distribution system, where such behavior is a direct result of a causal relationship between the internal and external elements of the system. The effects of causal relationships that are based on assumptions and decision rules then be formalized by using mathematical equations. System Dynamics focuses on the interrelationships into account non-linear with changing process that occurs throughout the time horizon, thus allowing the user to model very complex system [5]. With the System Dynamics framework, enable us to develop scenario and assessment for asset management strategic planning.

2 Method

System dynamics is a method to analyze and design a policy. Some steps is required to develop the model [6]: (a) problem definition which determine the significant variables; (b) model formulation which describe the relationship between components; (c) data collection; (d) model development; (e) model verification; (f) model validation used mean comparison and variance comparison [7]; (g) scenario development to improve the system performance (h) model interpretation; and (i) implementation.

3 Results

3.1 System Dynamic Model Development

Asset management planning is a strategic framework for asset management. It also shows how the company will serve the needs of the country's power sector that is efficient, reliable, and quality services. Impairment of assets conditions can substantially increase the overall maintenance cost, raises issues other risk management, and have a negative impact on the environment, the overall state economic and increase operating costs [8].

Asset management program support the objectives and performance targets. Network performance targets to be a reference for asset planning and regular maintenance in the next year. It is also a reference to the actions required; whether the asset should be replaced or just need the maintenance. In the end, the target is expected to improve customer satisfaction, as shown in Fig. 2. Network performance can be affected by several factors, including temperature, humidity, weather, geographic, and the usage [9].

The other factors are the lives of assets that have been set by the manufacturer also affect network performance. When the assets is getting closer to the limit of the age, the higher the frequency of damage to assets, as shown in Fig. 3.

Indications disorders are categorized into two types: the transmission and distribution. In distribution, the two indicators are often used of SAIDI and SAIFI. And in terms of transmission, two things became calculation that System Outage Duration (SOD) and System Outage Frequency (SOF) as shown in Fig. 4.

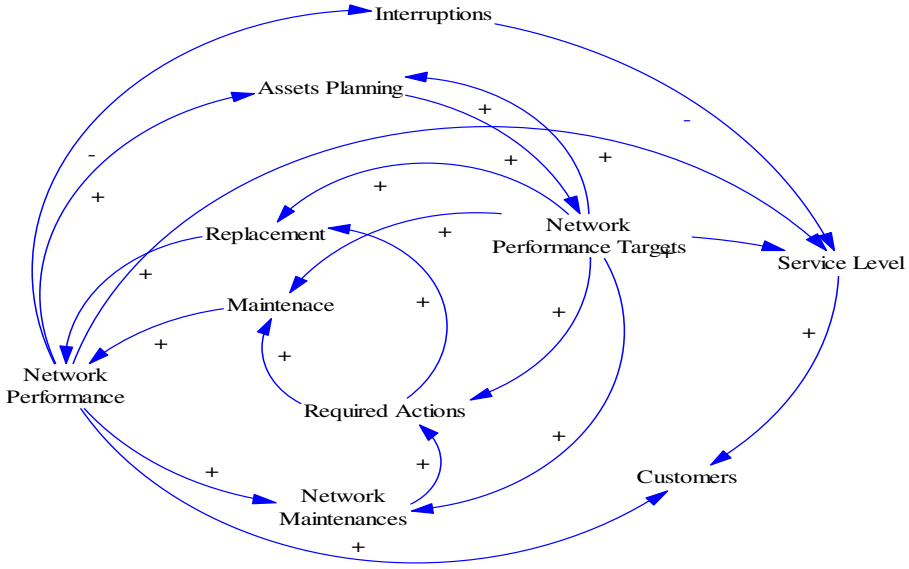


Fig. 2. The link between performance targets, maintenance, and service levels

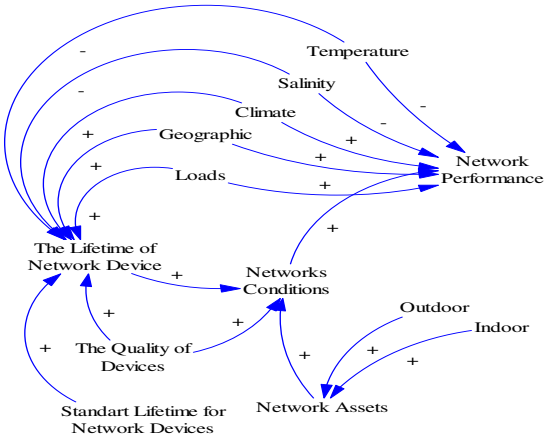


Fig. 3. The causal relationship of asset lifetime

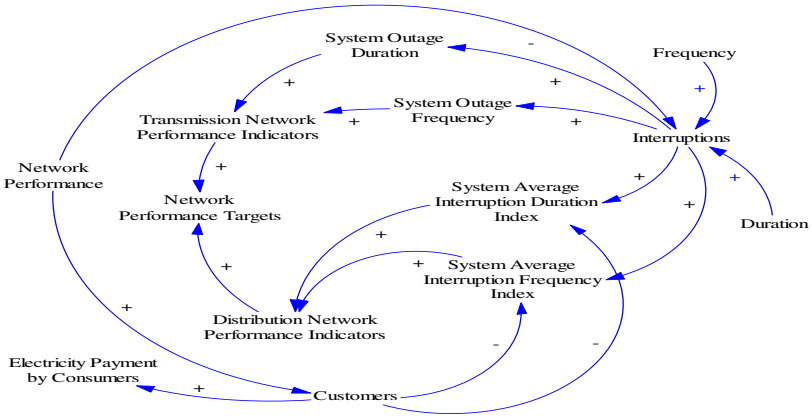


Fig. 4. The linkage of network performance indicators [8]

From the several significant variables above, caustics diagram can be developed as shown in Fig. 5.

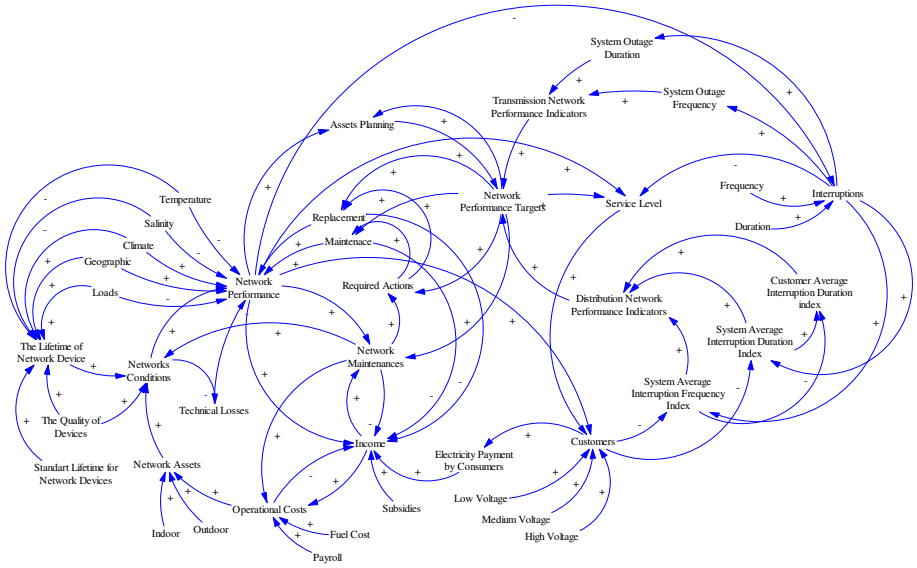


Fig. 5. Asset management distribution network caustic diagram

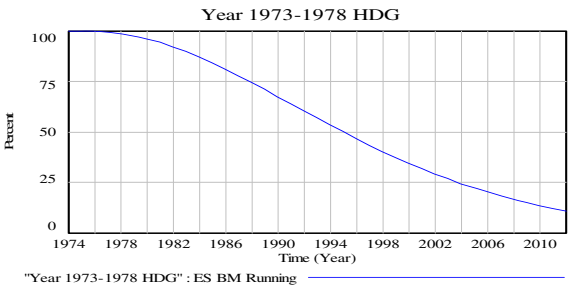
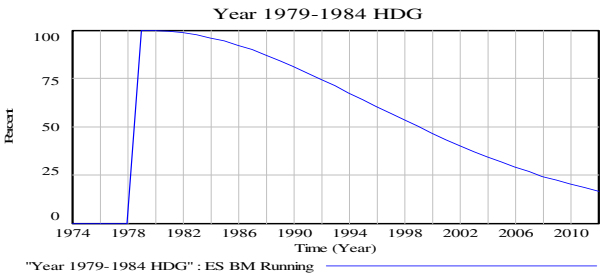
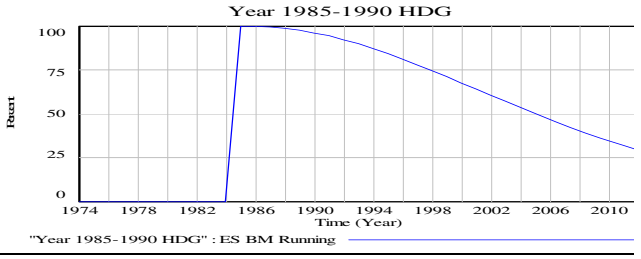
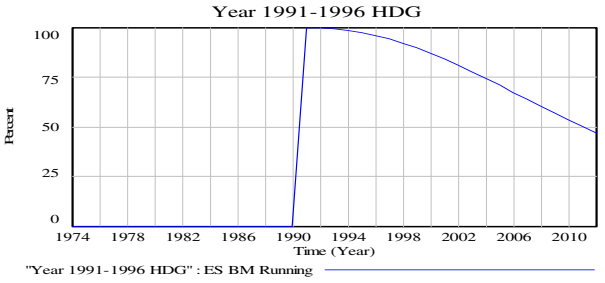
From Fig. 5, it can be seen that the asset management plan is determined by the network performance network performance targets. Network performance is affected by network conditions, asset maintenance, asset renewal, as well as external factors such as temperature, salinity, climate, geographical conditions, as well as network load. The target network performance is affected by the performance of the transmission network, the performance of the distribution network, as well as asset management planning. Service level is influenced by the frequency and duration of

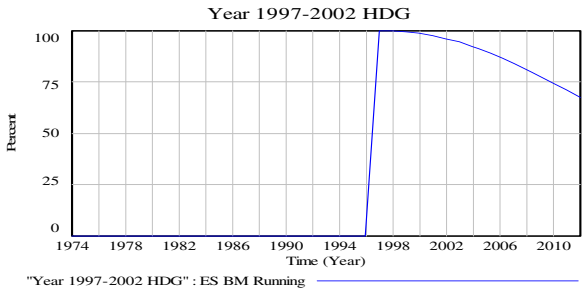
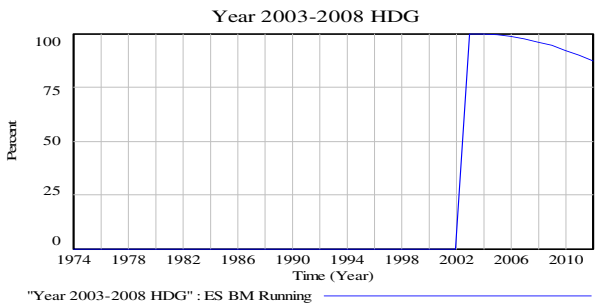
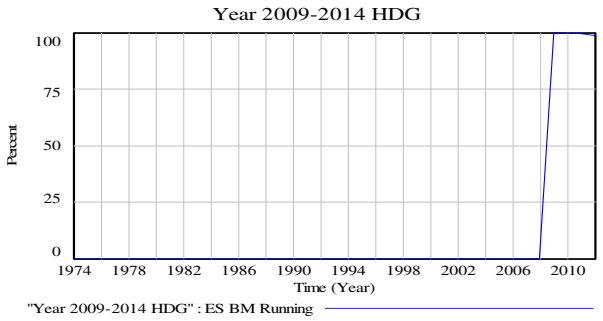
the disturbance (interruption), the performance of the network (network performance), as well as performance targets network to improve service levels.

3.2 Low Voltage Network Assets

From the simulation results can be seen that the low voltage network assets can be grouped into seven periods of the installation as shown in Table 1.

Table 1. Low voltage network asset condition by installation period

Year	Quantity	Low Voltage Network Asset Condition
1973-1978	3	
1979-1984	28	
1985-1990	387	
1991-1996	39.749	

Year	Quantity	Low Voltage Network Asset Condition
1997-2002	7.616	
2003-2008	21.750	
2009-2014	26.475	

From Table 1, it can be seen that the condition of the low voltage network assets installed in the period 1973 to 1990, has been under acceptable condition (<50%), so it is necessary to reform the asset. Overall, the average assets condition in the low voltage network by considering the assets installation can be seen in Fig. 6.

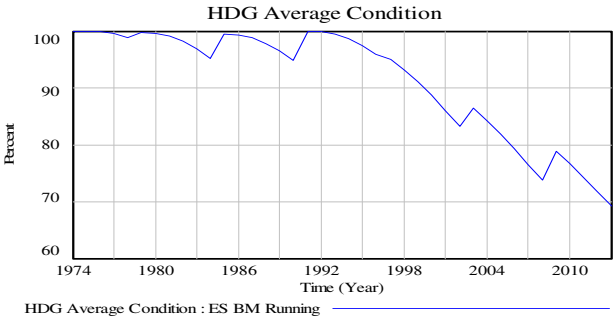


Fig. 6. The Average Assets Condition in the Low Voltage Network by Considering the Assets Installation

3.3 Transformer Assets

Based on data from PLN UPJ Pamekasan, transformers assets can be grouped into three installation periods as shown in Table 2. From Table 2, it can be seen that the condition of the asset transformers installed in the period 1979 to 1980, has been under acceptable condition (<50%), so it is necessary to reform the transformer assets. Overall, the condition of transformers on average assets by considering te installation of the assets can be seen in Fig. 7. From Fig. 7 it can be seen that the general condition of the asset transformers, already under acceptable condition, as a percentage of the bulk transformers which 72.95% had need of renewal (for the period 1979-1980).

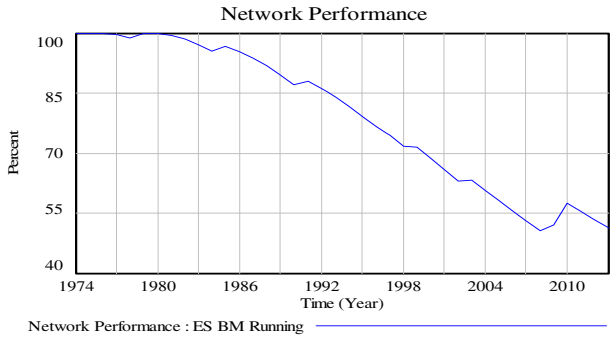


Fig. 7. Network Performance

3.4 Network Performance

The results of variable network performance can be seen in Fig. 8. While the results of variable performance losses as a result of the performance losses can be seen in Fig. 8.

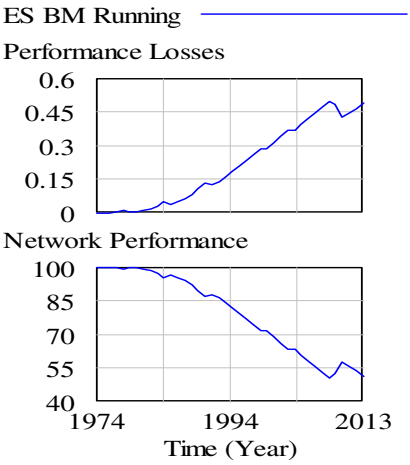


Fig. 8. Performance losses as an effect of network performance

3.5 Medium Voltage Assets

The results of the simulation shows that the average decline in the medium voltage network assets is approximately 4.4% per year as shown in Fig. 9. Conditions of medium voltage network assets installed in 1975-1990, under acceptable condition (50%), and in need of reform.

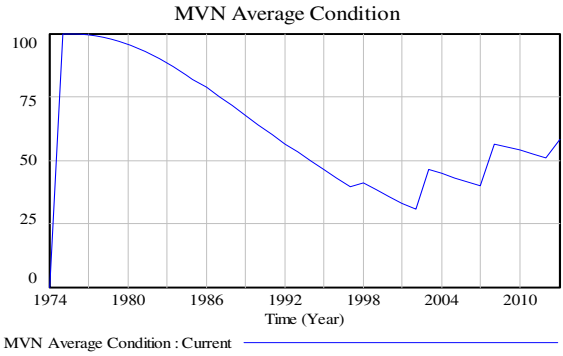


Fig. 9. Medium voltage assets

3.6 SAIDI and SAIFI

From Fig. 10 and Fig. 11, it can be seen that in the period 2003-2013, the average SAIDI = 7.4 hours / year, while the average SAIFI = 3.4 times / customers. SAIDI can be formulated as follows:

*SAIDI = (Duration Interruption * Number of Customer Impaired) / Total Customer* (1)

While SAIFI is the ratio of total customers who experienced outages on the number of subscribers that can be formulated as follows:

$$SAIFI = Number\ of\ Customer\ Impaired\ /\ Total\ Customer \quad (2)$$

The equation (1) and (2) are used to analyze data in the period 2003-2013.

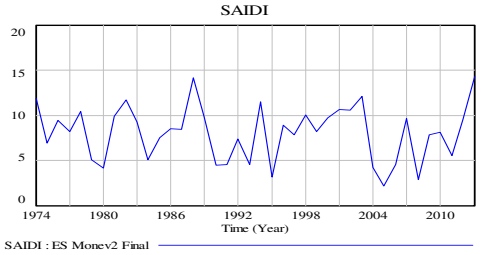


Fig. 10. SAIDI in UPJ Pamekasan

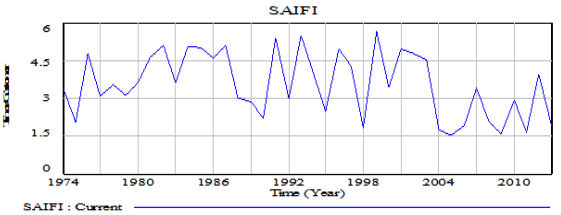


Fig. 11. SAIFI in UPJ Pamekasan

3.7 Technical Losses

The simulation results of monthly technical losses can be seen in Fig. 12.

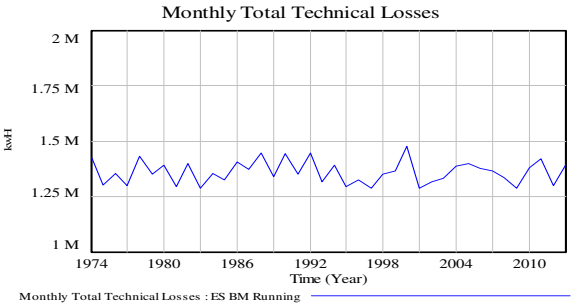


Fig. 12. Monthly technical losses

From Fig. 12, it is seen that the average technical losses in UPJ Pamekasan average of 1,359,981.60 KWH / month or about 16,319,779.24 KWH / year. Meanwhile, the comparison between data and simulation model of technical losses is given in Fig. 13.

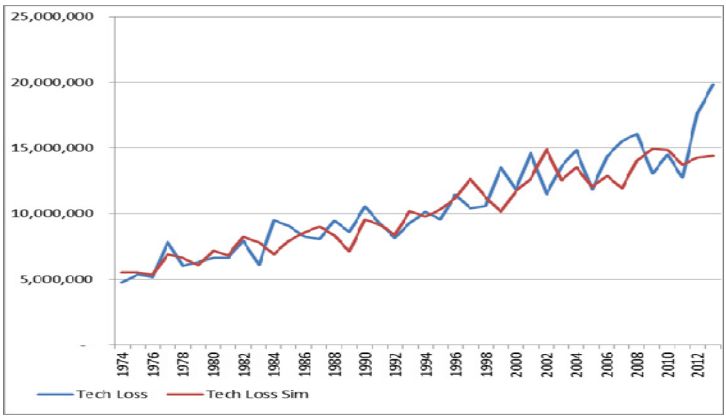


Fig. 13. The comparison between data and simulation model of technical losses

As we can see from Fig. 13, the Error variance is less than 30% (Error variance = 17.31%), which means the model is valid.

4 Conclusion

From the research that has been done, it can be concluded some important things such as following:

1. The original contribution in this paper is providing framework for developing model of assets management that accommodate the internal and internal factors that affect the assets condition, as well as link up the assets condition to SAIDI and SAIFI
2. Caustic diagram is required as a basic framework to develop a model.
3. The flow diagram illustrates the relationship between variables expressed in the form of symbols such as the level, rate, and auxiliary, so it can be used to develop a mathematical model.
4. In general, the asset condition is determined by the asset treatment (maintenance) and a decrease in assets (deteriorate), which can be influenced by internal factors such as the standard lifetime and external factors such as geographical conditions, the temperature, the quality of network components, network load, climate, and salinity.
5. Asset condition will decrease over time. It is caused by a decrease in asset condition (deteriorate) has greater effect, when compared with assets treatment (maintenance).

6. Decrease in low voltage network assets conditions, on average, in the range of 6% per year.
7. The average of decrease in medium voltage networks asset condition is in the range of 4.4% per year.
8. The average decrease in transformers the condition of, is approximately 6.6% per year.
9. Condition of the low voltage network is affected by the amount of total assets installed within a certain time and condition of the assets.
10. Conditions of transformer asset are affected by the total number of transformers installed in the certain period and conditions of the transformers.
11. The distribution network performance is affected by the overall condition of the property. Network performance will have an impact on the performance losses.
12. In the period 2003-2013, average SAIDI - average = 7.4 hours / year, while the average SAIFI in the year 1975 to 2013 = 3.4 times / customer.
13. The losses of electricity distribution network can occur in medium voltage networks, low voltage, house connections, and transformer losses.
14. The average technical losses in UPJ Pamekasan average of 1,359,981.60 KWH / month or approximately 16,319,779.24 KWH / year.

References

1. Top Energy: Network Loss Factor Methodology. Top Energy Limited, New Zealand (2008)
2. Schneider, J.: Asset Management Techniques. In: Proceeding of 15th Power Systems Computation Conference, Liège (2005)
3. Davidson, I.E.: Utility Asset Management in the Electrical Power Distribution Sector. In: Power Engineering Society Inaugural Conference and Exposition in Africa. IEEE (2005)
4. Gaul, A.J., Spitzer, H., Nilges, J.: Strategische Investitionsplanung - Praxisnahe Wege für eine nachhaltige Bewirtschaftung der Assets (Strategic Investment Planning - Practical ways for the sustainable management of assets). Energiewirtschaftliche Tagesfragen 54(10), 655–656 (2004)
5. Wenzler, I.: Development of an asset management strategy for a network utility company: Lessons from a dynamic business simulation approach. Simulation & Gaming 36(1), 75–90 (2005)
6. Sterman, J.D.: Business Dynamics. Systems Thinking and Modeling for a Complex World. McGraw-Hill/Irwin, New York (2000)
7. Barlas, Y.: Formal aspects of model validity and validation in system dynamics. System Dynamics Review. John Wiley & Sons, Ltd. (1996)
8. DIER: Strategic Asset Management Plan. Department of Infrastructure, Energy and Resources, Tasmania, Tasmania (2005)
9. Crisp, J.: Assets Man. Electricity transmission, Ph.D Dissertation (2004)

Enhancing the Student Engagement in an Introductory Programming: A Holistic Approach in Improving the Student Grade in the Informatics Department of the University of Surabaya

Budi Hartanto

Informatics Engineering, University of Surabaya, Surabaya, Indonesia
budi@ubaya.ac.id

Abstract. Student engagement has long been known can increase the student performance. However bring this concept to practice may not be as easy as it sounds. Some factors, such as the teacher, the students and their background, the course content, the academic atmosphere, the study culture, etc., influence its implementation. This paper presents the practice that is applied in the Informatics Department of the University of Surabaya to enhance the student engagement in the introductory programming course (i.e. Algorithm and Programming). This course is one of the course that is commonly known as difficult and make many Informatics students in the University of Surabaya dropped their study in the Department. The practice to enhance the student engagement in the Informatics Department of the University of Surabaya is designed to fit best with the condition in the Department. As a result, the students' performance increase two grade level compared to the students' performance in the previous years.

Keywords: Student engagement, introductory programming, increasing the student performance.

1 Introduction

Algorithm and Programming (Alpro) – an introductory to programming course in Informatics Department of University of Surabaya – has long been known as difficult and make many students in Informatics students in the University of Surabaya dropped their study in the Department [1, 2, 3, 4, 5]

Alpro is used as a prerequisite before the students can take the other programming course that is Object Oriented Programming (OOP). To take OOP, the students must pass Alpro with a minimum grade: C. In Semester I of academic year 2012/2013, 45% of the students failed to pass Alpro. In the following year – Semester I 2013/2014, the percentage of students who failed to pass Alpro is 44%.

The efforts to reduce the number of students who failed to pass Alpro were tried for several years. However these efforts tried to fix the problem from the cognitive point of view only. In this case the teachers tried to fix the problem of the student's

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 493–504, 2015.
DOI: 10.1007/978-3-662-46742-8_45

inability to read and understand student code [6], the difficulty in understand the syntax and semantics [7], etc. As a result, there were modification in the breadth and depth of course content, the level of difficulty of the assessment, the delivery method, and so on. Unfortunately, the number of students who failed to pass Alpro were still very high.

This research tries to find methods, approaches, or strategies that can be used to decrease the student failure in Alpro. A lot of research showed that the student engagement is a key point to improve the student performance [2, 3], [8, 9, 10]. Therefore increasing the student engagement seems as a promising method to solve the problem in Alpro. Unfortunately, the practical approaches and strategies to improve the student engagement may vary from case to case. Some factors such as the type of the course, the education culture, the student background, etc. may influence the approaches and strategies that can be used.

In this paper, some practical approaches and strategies that are used to increase the student engagement in Alpro, in the Informatics Department of the University of Surabaya are shown and discussed. These approaches and strategies are proven can effectively increase the students' engagement that in the end increase the students' performance. As a result, the numbers of students who failed to pass Alpro in Semester I, 2014/2015 is projected to become 12% only. This is 28% lower than the one in the previous year. In Addition, the average mark of the students is projected to increase two grade levels, from BC to AB.

2 The Historical Background of Algorithm and Programming

Alpro is offered in the Informatics Department of the University of Surabaya as an introduction to programming course. This course is designed to be administered in a classroom for four unit hours (1 unit hour = 50 minutes) per week. In addition, the students can also take two more unit hours (optional) in the computer lab. A teacher administers the meeting in the classroom, while a student assistant administers the meeting in the lab.

The meetings in classroom were intended to give a conceptual view of programming to the students. In some meetings the students were also asked to do some programming exercises. The solutions of these exercises can be written in the book. The intention to use classroom instead of lab was meant to increase the number of exercises that could be solved by the students. This was possible because the students do need to run and debug the program that they wrote. Sometimes, finding errors from even a logically correct program can spend a lot of time.

In this scenario, the students were expected to copy the solutions in the book to the development software in their computer, at home. Unfortunately, not many students performed this. As a result, it was common to find students that never use the development software. Some of them even did not have the software installed in their computer.

The existence of the lab session did not help too much. Because the students' attendance were not compulsory, the number of students in the lab session dwindle quickly week by week. The lab session that was meant to help them in the programming problem

and to support them to understand programming better, did not fascinate the majority of the students.

As described in the Introduction section, 45% of the students failed to pass Alpro in Semester I 2012/2013, and 44% in Semester I 2013/2014. The students are considered failed to pass Alpro if their grade is D or E.

3 Holistic Approach to Enhance the Student Engagement

Theoretical and empirical research suggests that academic engagement can improve the student performance [2, 3], [8, 9, 10]. However the practical strategies to improve the student engagement may different from case to case. A holistic approach may be required to truly actualize the student engagement in class.

In the following subsection I will discuss some factors that influence the student engagement, as well as the approaches and strategies that I and the teaching team perform to improve the student engagement.

3.1 The Course Content

Learning programming is a skill [7]. Human commonly masters a skill by performing hands on experience on the skill to be mastered. In learning how to program, the hands on experience are writing the program and learning how to find and fix the error.

That is the reason why teachers always ask students to write the program for solving the task in the exercise. Unfortunately, experience from the previous semesters showed that many students were not motivated to write the program. The engagement of the students to the class activity were very low.

One contributed factor to this condition is the content of the course that do not support them to understand well and cannot arouse their excitement. In order to increase the student engagement from the course content factor, we perform the following approaches and strategies:

- Problem based learning
  One reason that brings the students disinterest in writing programs is the absence of clear relation between the problem in the exercise to the problem that they thought they would face in real life [11]. For example, one common exercise in Alpro is asking the student to write a program to find factorial numbers. The majority of the students may thought that the program to find factorial numbers was meant as a programming exercise only. They thought that they will never face this kind of problem when they work as programmers later on. This kind of thought will certainly decrease the students' motivation to write the program.

  Our strategy for this case is to rewrite all the given exercises in the previous semesters, and make them similar to the problem that the students thought they may face in real life. By this kind of exercise, we expect that the exercise would fit better to the student expectation if they work as a programmer later on.

As an example, to ask the students to write a factorial program, we create a story of a certain matchmaking company that wants to hire a programmer to solve a certain problem. In this case the company problem is to find out the number of possible combination to match the men and women that are already registered in that company. Certainly the problem of finding the number of combination can be solved by finding a factorial number. Applying the knowledge to be learnt to the real life problems can increase the student interest in learning [12].

- Building on prior knowledge
  It is important to know that many students in the Informatics Department of University of Surabaya have no background in programming. Introducing programming as a new and isolated entity will make it hard to be understood or to be retained in the students' memory. Human memory works by connecting the new information to the existing information that they had in memory [13].

  Because the students may not have a programming background then we need to link the concept of programming to the events that they have experienced in their life. We use analogy. For example, to introduce the concept of how a program works, we gave an analogy of cooking an instant noodle. When a person cooks an instant noodle using the exact steps as written on the noodle package, the detail of the instructions, the usage of the appropriate words, and the correct order of the instruction will influence the result that the person gets. Using this analogy, the students are expected can imagine the concept of programming.

  In addition, in explaining the new topic we tries to link it to the previous topics. Learning can be effective if learners can connect the new material with the one that already exists in memory [14]. Linking the new topic to the previous topic can make the new topic become easier to be understood and also strengthen the knowledge of the previous topic.

- Learning by example
  Human learn by example [15]. Therefore it is important to provide an example and let the students do the exercises by mimicking the work in the example. In previous semesters, examples were also given in the end of certain topics. Unfortunately, after further analysis we found that some examples were stressing on how to use the program constructs rather than solving a problem using the program constructs. An example that shows how to solve a problem is more beneficial than an example that shows how to use the program construct.

  We checked all examples in the previous semesters and revised them to become examples that show how to solve certain problems. We ensure that the first exercise has a similar solution to the example solution but it is wrapped in different case study. Therefore, the work in the example can be used as a foundation of thought to do the first exercise. In addition, we also ensure that all topics are always ended with an example.

- Learning by doing
  Because programming is a skill [7], the students should have hands on experience of it. Learning how to program can be mastered if the students often practice writing the correct program [12, 13]. Unfortunately experience from the previous semesters showed that not many students were motivated enough to write the program, especially in the development software.

  In this research we applied some strategies to encourage and to push the students to write the correct program in the development software. The first strategy that we performed was asking the student to install the development software in their computer. The development software that we use is Visual Studio Express Version that can be downloaded and used for free. To motivate the students, we rewarded a 10 point mark to the students that can show that Visual Studio are already installed in their computer. Some students sent a screen capture of the software and some students were happily sent their selfie photo with Visual Studio displayed on their monitor. The good thing is that all students did this task.

  The existence of the development software in the student computer is the first step required to encourage the student to write the program in the development software. At any time the student wants to write the program, the development software is already in their computer, ready to be used. On the contrary, without the development software ready, the students that originally want to write a program, must spend some time to download and install the development software. The lengthy process of downloading and installing the software may put out the students intention to write the program.

  The second strategy is to encourage the students to bring their laptop to the classroom. With this laptop, the students can write the exercise solution directly on the development software. They do not need to write the solution in the book first and copy it to the development software later on. Because of this encouragement and other strategies from other factors, there are many students bring their laptop to the classroom in this semester. Writing the program on the development software enables the students to immediately see the result of their program. A program that can run and display the result correctly often give a certain level of satisfaction to the programmer.

  If the students did not bring laptop, we still wanted the students to write the program in their book. Therefore we repeatedly reminded the students that learning programming is a skill and can only be mastered by practice writing programs. We told the students that we would really happy if they can show their program to us to get any feedback. Many students were happily showed their program to us for a feedback. The teacher feedback is one of the situational factors that affect student learning [17].

  The third strategy is to balance the time used to deliver the knowledge and to practice writing the program. We believe that the students need to practice writing the

program, but we also believe that they need to understand the knowledge that underlie the practice. Therefore learning the knowledge and practice is equally important [18] Because of this we strive to use only 50% of the allocated time in every meeting to deliver the knowledge and to use another 50% for practice.

Although this strategy looks simple, its implementation was not easy. In the previous semesters, one topic was usually delivered as a whole (just like a book chapter) before it was continued with the exercise. Because of this strategy, we have to split a certain topic into several sub topics. We have to ensure that the sub topic can be delivered in more or less 50% of the allocated time. The application of this strategy can successfully make the students understand the topic better.

3.2 Academic Atmosphere

In previous semester, it was common to see a passive and apathy students in class. These students preferred to use their time for chatting with their friends instead of doing the exercises. As a result, the students that actually wanted to do the academic activity eventually got affected and joined the chat. The engagement to the class activity was low. The reasons that make the students become passive and apathy are possibly from the students' disability to do the exercise but they are too shy or afraid to ask the teacher about how to do it. This is a common academic problem in many education level in Indonesia. The following are some strategies that we do to improve the academic atmosphere.

- Encourage collaborative work

Collaborative work or peer learning is known to be good to enhance the student engagement [19, 20]. In this semester, we encourage the students to perform collaborative work. They can informally make a group and have discussion to find the solution of the exercise. From our observation we can see that in every group there are always one or two students who look smarter than the others. These students were not reluctant to share their knowledge to the less-able students when they asked for help. The students were not shy or afraid to ask questions anymore because now they could asked the questions to their friends. No more reason for the less-able students to say that they do not know how to solve the problem because they can discuss the solution of the problem with their friends.

- Encourage a good student-teacher communication

A good communication between the student and the teacher is very crucial to make the student become more open to the teacher. It is well known that Indonesian students are very shy and very reluctant to ask to the teacher. From my own experience as a student and from my observation as a teacher, this behavior was possibly formed due to the traumatic experience of asking questions in class. It is common for Indonesian students or even teachers to give funny comments about the questions or about the person that asks the question. Therefore, many students prefer not to raise questions to avoid the embarrassment.

In order to achieve a good student-teacher communication, we try to get closer to the students by going around the classroom at the exercise time. Therefore the students do not need to be afraid to be mocked by their classmates because now they can ask us personally and no need to say it loud in front of their classmates. In addition, the other strategy that we did – that encourage the students to show their work to us – align with this strategy. We performed a verbal and nonverbal immediacy when the student showed his/her work to us. Verbal and nonverbal immediacy support a better learning [21]. This strategy is proven can make a good student-teacher communication and can make the student become more open to the teacher.

3.3 The Teacher

Another factor that influence the student engagement is the teacher role in class. Some approaches and strategies are applied in this semester to increase the student engagement from this factor.

Be a Motivator, not Just a Provider. In many classes, it is very common to see teachers spend all of their teaching time to deliver and explain the knowledge to the students. This behavior come from the common thought that the teachers' job is to deliver knowledge to the student. Not many teachers realize that it is also their responsibility to motivate the students and arouse the students' interests.

With the current advance of the searching engine, explaining the full features of a topic are not critical anymore. The teacher can just explain the most common and important part of the knowledge and ask the students to explore deeper by searching from the internet search engine. By this approach the teachers can have some spare time that can be used to motivate the students or to arouse the students' interest. This is especially needed for students with a weak background in the subject [22]. A motivated student would willing to enrich themselves with a deeper knowledge.

Give Feedback, not Only Judgment. Assessment can be classified into two categories: formative and summative. The formative assessment is used to gather information for the specific purpose of improving learning (for learning), while the summative assessment is used to gather information for the purpose of recording and reporting (of learning) [23, 24]. In formative assessment we can get the feedback of the current teaching and learning condition and use the feedback to improve them. On the other hand, the summative intention is solely to form a judgment [25].

There are six assessments that are performed for the students in Alpro. These assessment are administered in Week 4, 6, 8, 12, 14, and 17. In the University of Surabaya, the teaching and exam weeks have a pattern of 7, 2, 7, 2. The first seven weeks are used as the teaching weeks and the next two weeks are used as the exam weeks. This pattern is repeated again after the exam weeks. The teacher can perform any assessment in the teaching weeks but cannot perform a lecture in the exam weeks.

Although all assessment marks contribute to the students final grades, five of them actually can be considered as formative assessments. This is because the result of the assessments can be used to give feedback to the teacher and the student about the

current condition of the teaching and learning. The sixth test is performed in the second exam week and is used solely to measure or judge the student performance.

The number of students in each Alpro class is around 40 to 50. If for example the total number of students in a particular semester is 300 then there will be 6 classes open for Alpro. The number of teacher is usually set to be the same as the number of classes.

In order to avoid the marking variant between teachers we usually assign one teacher to mark only one questions in a particular assessment but he/she should mark the work of the students in all classes. If the number of students is 300 for example, then one teacher will mark 300 students' works. Checking the works of 300 students in a short time is challenging and need a big commitment to finish it before the next test. Returning the students' work at or after the next test will make the formative assessment lose its essence as a feedback instrument. Unfortunately, in the previous semesters due to the tight schedule of the assessments, teachers often return the assessment result at or after the next assessment.

In this semester, we make a commitment to return the assessment result at most one week after the assessment. With this commitment, the formative assessments can now be used to give a real feedback to the teacher and the student about their performance. As described by [26], an assessment can become catalyst to the student performance, the validity of the program, and the validity of the student performance. A late returned assessment can only benefit in the validity of the student performance.

3.4 The Students

One last factor in our consideration to improve the students' engagement is the students themselves. Some approaches and strategies that are applied in this semester from the students' factors point of view are described below.

- Increase the student self-efficacy

Many ordinary students lack of belief that they can solve the given problem with their current capability (lack of self-efficacy). Self-efficacy itself is defined as beliefs that build a person's motivation and confidence in achieving a desired outcome [27, 28]. Without self-efficacy the students will give up too early and not use their best efforts to find the solution of the problems.

There are several strategies that we perform to increase the student self-efficacy. As described by [17], the students will have a high self-efficacy if the students can solve difficult problem or do not spend too much effort to solve a problem. On the contrary, the students that solve an easy problem or spend too much effort to solve a problem will have a low self-efficacy. Therefore our strategy to increase the students' self-efficacy is wrapping some easy problems with applied cases to make them appear like advance problems. As an example, the problem of finding a factorial number is wrapped in an applied case such as finding the number of possible combination to pair man and woman in a matchmaking company. This wrapper will make the original problem of finding a factorial number looks more advance. However, finding the solution for this problem is not hard.

In another strategy to increase the student self-efficacy, we create an exercise where its solution is similar to the example given. However we wrap this exercise in different case study to make it looks different from the case study in the example. As a beginner programmer, the students usually do not realize that the solution for the exercise is actually very similar to the one in the example. We expect the student to consider the difficulty of the exercise is at medium level (not at easy level) and we also expect the students can solve the problem without too much effort.

- Provide help as needed

Helping students in a medium or a big size class can be problematic. On one hand we want the students to ask question if they have problem, but on the other hand we know that we will not be able to answer many of the students' questions if many of them raise questions at the same time. The inability to answer the students' question may frustrate the students and demotivate them to learn.

The strategy to overcome this problem is related to the strategy in section 3.2. In that strategy we encourage the student to perform collaboration learning so they can help each other. With this collaborative learning the students only ask questions to the teacher when all students in the group do not know the answer. An effective collaborative learning, helps reducing the number of questions significantly.

Although the collaborative learning can reduce the number of questions, it is possible that there are still many questions raised to the teacher at the same time. In the programming class, the condition can be worse. The correct program for a given exercise can vary from student to student. A good teacher should be able to see the student intention in his/her program and give help according to the current student program. The process to understand the student program itself can take a considerable time before another time is needed to think how to align the student intention with the correct solution and explain the solution to the student. Only experience teacher in programming that can performed these process quickly. Based on this need, our strategy to provide help to as many students as possible is by assigning a senior and experience teacher in programming [5].

4 Evaluation

The approaches and strategies described in here are already applied to the Alpro class in Semester I 2014/2015. As a result, the student engagement in class is improved a lot. The students become active in every activity given and the academic atmosphere in class is very positive.

The data in table 1 shows the percentage of students that failed Alpro in Semester I 2012/2013 and Semester I 2013/2014. Based on the students' whole marks (marks in mid and final semester), 45% students failed to pass Alpro in year 2012/2013 and 44% in 2013/2014. This percentage is similar to the percentage of the students who failed to pass Alpro if we used their mid semester marks only to judge their grade (41% and 40%). The actual number of students that failed to pass Alpro is 4% worse than the number of students that failed to pass Alpro based on their mid semester marks only.

Table 1. The comparison of the student failure and the student mid semester marks

	Number of students	Student failure		Average marks	
		based on the mid semester marks	based on the whole marks	based on the mid semester marks	based on the whole marks
2012/2013	193	41%	45%	59.62	54.29
2013/2014	245	40%	44%	61.06	57.34
2014/2015	255	12%	-	75.29	-

Therefore based on the students' marks in the mid semester, we can estimate and project the percentage of the students that will actually failed to pass Alpro in the end semester. When this paper is written, the students are still in the end of the teaching week and have not completed their final exam. Therefore we cannot get their whole marks yet. Because of this we will compare the student performance across academic years by using the mid semester marks to calculate the students' grade.

Based on the data in table 1, we can see that the students' failure in Semester I 2014/2015 decreases significantly to the students' failure in the previous academic years. If the pattern of the students' failure in academic year 2014/2015 is the same as the one in the two previous years, then only 16% (12% + 4%) of the students will fail in the end of semester I 2014/2015. It means that the number of students' failure in this semester decreases 28% to the students' failure in the previous semesters.

The average of the students' performance based on their mid semester marks also increase significantly. The students' average marks in year 2012/2013 and 2013/2014 are 59.62 and 61.06. In year 2014/2015 the average marks increase to 75.29. If we convert this values to grades then the students' average marks have increased from BC to AB. Therefore the students' performance is projected to increase two grade level higher (the grades in the University of Surabaya are: A, AB, B, BC, C, D, and E)

5 Conclusion

Improving the student engagement is proven can increase the student performance. However improving the student engagement can be different from case to case and must be elaborated into practical approach and strategies. These approach and strategies must be aligned with the teaching pedagogies, the type of the course, the education culture, and the student background.

This research shows that a holistic approach and strategies to improve the student engagement work well in an introduction to programming course named Alpro, in Informatics Department of the University of Surabaya, in Semester I 2014/2015. As a result of applying the approaches and strategies described in here, the students are willing to participate in the class activity and some of them are even keen to complete their work even though the class is over already. This students' engagement are proven can increase the student performance. Based on the mid semester marks, the students' failure in the end of semester is projected to decline to 16% (12% + 4%). This projected

students' failure is 28% smaller than the students' failure in the two previous academic years. The average students' performance based on their mid semester marks also shows an improvement. The students' performance in Semester I 2014/2015 is projected to increase to two grade level higher than the ones in the previous two academic years.

References

1. Kinnunen, P., Malmi, L.: Why students drop out CS1 course? In: Proceedings of the Second International Workshop on Computing Education Research, pp. 97–108. ACM (2006)
2. Woods, E.G.: Reducing the dropout rate. School Improvement Research Series (SIRS): Research You Can Use. Northwest Regional Educational Laboratory, Portland (1995)
3. Kushman, J.W., Sieber, C.: Heariold - Kinney, P.: This isn't the place for me: School dropout. In: Capuzzi, D., Gross, D.R. (eds.) Youth at Risk: A Prevention Resource for Counselors, Teachers, and Parents, 3rd edn., pp. 471–507. American Counseling Association, Alexandria (2000)
4. Corney, M., Teague, D., Thomas, R.N.: Engaging students in programming. In: Proceedings of the Twelfth Australasian Conference on Computing Education, vol. 103, pp. 63–72. Australian Computer Society, Inc. (2010)
5. Beaubouef, T., Mason, J.: Why the high attrition rate for computer science students: some thoughts and observations. ACM SIGCSE Bulletin 37(2), 103–106 (2005)
6. Mannila, L.: Progress reports and novices' understanding of program code. In: Proceedings of the 6th Baltic Sea Conference on Computing Education Research: Koli Calling 2006, pp. 27–31. ACM (2006)
7. Robins, A., Rountree, J., Rountree, N.: Learning and teaching programming: a review and discussion. Journal of Computer Science Education 13(2), 137–172 (2003)
8. Dev, P.C.: Intrinsic motivation and academic achievement: What does their relationship imply for the classroom teacher? Remedial and Special Education 18(1), 12–19 (1997)
9. Martin, A.J.: Examining a multidimensional model of student motivation and engagement using a construct validation approach. British Journal of Educational Psychology 77(2), 413–440 (2007)
10. Afzal, H., Ali, I., Khan, M.A., Hamid, K.: A study of university students' motivation and its relationship with their academic performance. International Journal of Business and Management 5(4), P80 (2010)
11. Kurkovsky, S.: Mobile game development: improving student engagement and motivation in introductory computing courses. Computer Science Education 23(2), 138–157 (2013)
12. Hartanto, B., Reye, J.: Incorporating Anchored Learning in a C# Intelligent Tutoring System. In: ICCE 2013, p. 5 (2013)
13. Mohs, R.: How human memory works (2007), `http://science.howstuffworks.com/life/inside-the-mind/human-brain/human-memory2.htm`
14. Mayer, R.E.: The psychology of how novices learn computer programming. ACM Computing Surveys (CSUR) 13(1), 121–141 (1981)
15. Robinson, A.: Children learn by example (1998), `http://www.home-school.com/Articles/children-learn-by-example.php`
16. Weragama, D., Reye, J.: Analysing Student Programs in the PHP Intelligent Tutoring System. International Journal of Artificial Intelligence in Education 24(2), 162–188 (2014)
17. Schunk, D.H.: Self-efficacy and cognitive skill learning. Research on Motivation in Education 3, 13–44 (1989)

18. Eckerdal, A.: Novice programming students' learning of concepts and practise. Ph.D. dissertation, Uppsala University, Division of Scientific Computing, Numerical Analysis (2009)
19. Jacobson, N., Schaefer, S.K.: Pair programming in CS1: overcoming objections to its adoption. ACM SIGCSE Bulletin 40(2), 93–96 (2008)
20. McDowell, C., Hanks, B., Werner, L.: Experimenting with pair programming in the classroom. ACM SIGCSE Bulletin 35(3), 60–64 (2003)
21. Witt, P.L., Wheeless, L.R.: An experimental study of teachers' verbal and nonverbal immediacy and students' affective and cognitive learning. Communication Education 50(4), 327–342 (2001)
22. Beheshti, M., Aló, R.A., Fernandez, J., Gates, A.Q., Ranjan, D., Boadi, A., Barker, L.: Work in progress-CS0 course implementation in Computer Science. In: 38th Annual Frontiers in Education Conference, FIE 2008, pp. F4A-7. IEEE (2008)
23. Taras, M.: Assessment – summative and formative – some theoretical reflections. British Journal of Educational Studies 53(4), 466–478 (2005)
24. Taras, M.: Summative and formative assessment Perceptions and realities. Active Learning in Higher Education 9(2), 172–192 (2008)
25. Wiliam, D.: Formative assessment: Getting the focus right. Educational Assessment 11(3-4), 283–289 (2006)
26. Martin, J.B., Collins, R.A.: Formative and Summative Evaluation in the Assessment of Adult Learning. In: Assessing and Evaluating Adult Learning in Career and Technical Education, 127 (2011)
27. Collins, J.W., O'Brien, N.P. (eds.): The Greenwood dictionary of education. ABC-CLIO (2011)
28. Bandura, A.: Self-efficacy: The exercise of self-control. WH Freeman & Co., Gordonsville (1997)

Business Process Maturity at Agricultural Commodities Company

Lily Puspa Dewi, Adi Wibowo, and Andre Leander

Informatics Department-Faculty of Industrial Technology-Petra Christian University, Surabaya, Indonesia
{lily,adiw}@petra.ac.id, leeshienwen@gmail.com

Abstract. The agricultural commodities company nowadays strive to keep transforming their business processes in accordance with the fast changing demands to survive the intense global competition. In an attempt to provide stakeholder with an insight business process, this paper investigates how to model business processes with Business Process Modelling Notation and assesses the process maturity using Gartner's model. This research is based on an in-depth observation at purchasing and production division of agricultural commodities company. It is found the business process model presents the knowledge from business in existing context. The other findings show the maturity level of the people is on phase 1, the maturity level of IT achieves phase 3, and the maturity level of the other four factors including strategic alignment, culture and leadership, governance and methods have already arrived phase 2. These findings will help company to do big planning for improvement in the future.

Keywords: agricultural company, business model, process maturity.

1 Introduction

1.1 Background

Agriculture can be one of the most satisfying and rewarding ways to make a living. From farming and food production to agribusinesses contribute to the health of the population and food security of the country. The agricultural commodities companies nowadays strive to keep transforming their business processes in accordance with the fast changing demands so as to survive the intense global competition. Creating a viable value-added agriculture business involves finding an idea or opportunity in the marketplace and then building a viable business to take advantage of the opportunity. This means improving business process is important to be done continuously in many aspects [1]. Therefore, business process management (BPM) becomes the most popular business and technology management method recently [2]. When company has carried out BPM, they have to know where and how the improvement can be done. Gartner introduce the BPM maturity model to measure the company's maturity phase

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 505–513, 2015.
DOI: 10.1007/978-3-662-46742-8_46

from many aspects. With Gartner's model, the company will know what phase they are and which factors should be re engineered.

1.2 Research Goal

This research assesses the agricultural business process and evaluates its maturity using Gartner model. Garner model is used because this model presents information to identify characters of each maturity phase. Gartner model focuses on six factors which are strategy alignment, culture and leadership, people, governance, methods and information technology. These six factors is called six critical success factor. Each factors have different maturity phase.

2 Literature Review

Business process is a group of related task performed to reach a specific business outcome [3]. Business services today have become increasingly active, knowledge-intensive. Therefore, it is quite natural that method to describe the existing business. In this case, the document flow systems are much more flexible and better adapted to the behavior of the business [4].

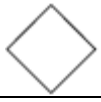
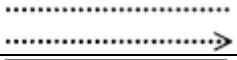
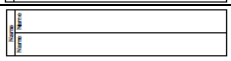
2.1 Business Process Management (BPM)

The concept of BPM is mainly comes from Business Process Reengineering (BPR) [5]. The main focus of BPR is business process and information technology, and its success is built on the good communication and collaboration of information technology and business [6]. With the IT support such as hardware, software tools and network, BPM helps to achieve the optimized and value creation of operational business processes [7]. Other researcher said that BPM is IT-focused which characteristics BPM from the perspective of business process automation [8].

2.2 Business Process Modeling Notation

Business Process (BP) models constitute a graphical representation of processes in an organization. Business Process Model and Notation (BPMN) [9] is a notation for modeling Business Processes, which contributed significantly in Software Engineering when it comes to collaboration between developers, software architects and business analysts. Business Process Modeling Notation (BPMN) is a notation that describes the logical steps in the business process [10]. This notation has been especially designed to coordinate the sequence of processes and messages that flow between participants in different business activity. The notation of BPMN can be seen at Table 1.

Table 1. BPMN modeling elements. (Source: Business Process Modelling Notation, p. 18)

Element	Description	Notation
Event	Event indicate that something is happening during the business process. Events affecting the flow and is usually caused by a trigger or an impact as a result of the event,	
Activity	Activity is a notation that describes their work done. Activity can be atomic (one event) or Sub-Process (set of activities)	
Gateway	Gateway serves as a control in order to split or merge process flow.	
Sequence Flow	Sequence flow is used to display the sequence of activities contained in a process.	
Message Flow	Message flow is used to show the flow of messages between the two participant (entity) the sender and recipient. In BPMN, the existence of two separate Pool indicate a second participant.	
Association	Association used to provide information regarding data object involved	
Pool	Pool describes participant in a process. Usually illustrates B2B situation.	
Lane	Lane is part of the Pool which usually categorize the activity / functional split participant.	
Data Object	Data objects are interpreted as an artifact that has the information.	
Group	Grouping is used to categorize activities possess similarities category.	
Text Annotation	Text Annotation present additional information in the BPMN diagram.	

2.3 Gartner Maturity Model

Based on Gartner maturity model, it has six phases BPM maturity model that involves understanding the six phases of BPM maturity and where the company stands on addressing critical success factors defined in a BPM maturity framework [11]. The six phases of BPM maturity can be seen at Fig. 1.

The six phases begin in **Phase 0**: company realizes that some business improvement cannot be done by traditional approaches. The company needs to seek the causes of bad performance. In **Phase 1**, company realizes "process aware" and take some actions to improve the process. In **Phase 2**, company starts to do the business process with IT support to gain better control in internal company. In **Phase 3**, the company build the relationship between internal company and external company (customer and supplier), cross-department cooperation and integration become a habit gradually. In contrast to phase 3, in the **Phase 4**, organization will link strategic goals to process execution directly. Ultimately, in the **Phase 5**, the company will do the agile business structure, that helps to capture new opportunities, gain more profits, knowledge about market demand, resources, competition and partners [11].

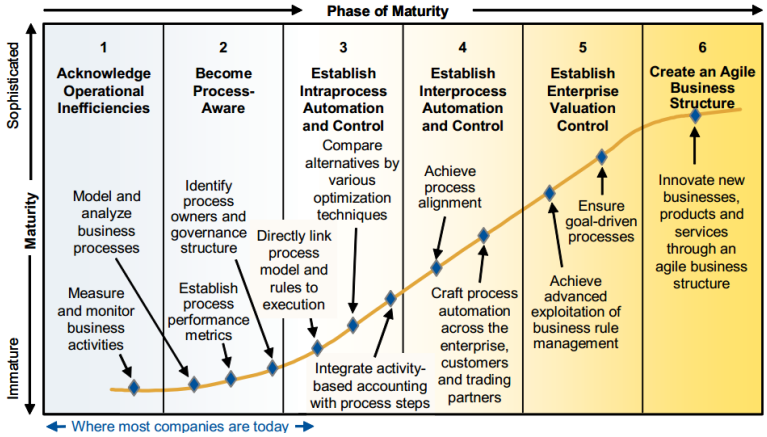


Fig. 1. The six phases of BPM maturity. (Source: Gartner, p. 4)

2.4 Six Critical Success Factors

The important dimension is the organizational factors must be balanced in maturity phases called critical success factors [11]. The six success factors are [11, 12]:

- Strategic alignment: the linkages that connect that connect organizational priorities and enterprise processes in order to achieve its business goals.
- Culture and leadership: culture is the values, beliefs, attitudes and behaviors that presents the company identification and public images. Leadership is crucial roles in supporting process design and facilitating cultural change throughout the company.
- People: The individuals and groups who continually enhance and apply their knowledge and professional skill related company's processes.
- Governance: refer to the transparent accountability, responsibility, decision making and reward processes to guide actions.
- Methods: The approaches and techniques to support consistent process for success of company.
- Information technology: The software, hardware and information management systems that enable and support process activities.

3 Research Methodology

The main methods used is qualitative research. The detailed and specific data can be collected through observing, interviewing and document. This research assesses the BPM maturity level in an agricultural through the critical success factors including strategic alignment, culture and leadership, people, governance, methods and information technology. The data collection is done with the primary data (interviews) and secondary data (documents).

In some qualitative research studies, the data and/or findings are returned to participants in order to obtain their validation. In this research used member validation or

member check. Member check is a procedure largely associated with qualitative research whereby a researcher submits materials relevant to an investigation for checking by the people who were the source of those materials [13].

4 Analysis and Results

4.1 Business Process Modelling

Business in agricultural sector has special business processes. The business process focused in this research is the agricultural company for purchasing and production division.

Purchasing Process. In this case, farmers as the supplier to this company could apply loans for farming capital. Farmers do a contract for selling their agricultural products with the company. This contract contains the number and agricultural products, due date, transportation costs and dealing price. Usually, the contract is made in couple of days before delivery time. On due date, farmers will bring agricultural products to the company warehouse directly, but in some cases the company may take the agricultural products from the boats as a delivery transportation. Furthermore, these agricultural products are weighed and samples taken for quality control such as water content, shell, gems, and rootlets. The quality control result determines the price. The beans' water content is a very important factor and should be 6 to 8%. If the water content is too high, microorganisms will spread and their quality will decrease. This water content determine the quality of the agricultural product, furthermore, it affects to the price.

In some cases, the farmers may entrust their agricultural products to the company for the higher future prices. That means, the agricultural products is already weighed, but the company does not give the dealing prices. The company provides this opportunity only for farmers who have become regulars or carry items with a large amount. The aim is that the seller can get a higher price at a later date when prices rise.

The next steps after weighing products, the warehouse clerk will send the invoice to the finance. Finance will check the quantity and price. If the farmers has a contract, the price will be same as stated the contract. The price also depend in the quantity of products.

Production Process. Agricultural products will be dried to reduce the water content. The beans' water content is a very important factor and should be 6 to 8%. If the water content is too high, microorganisms can infest the beans and their quality will be decreased. This drying process do in two ways, first is roasting whish roaster machine and second is still do in traditional way such as the beans is dried under the sun. The traditional drying process is do because the lack of roaster machine meanwhile the climate support this process. The next step is packaging. The agricultural product will be packed in sack with average weight of 70kg. After that, the sack will be put in the warehouse.

The business process modeling for purchasing is shown at the Fig. 2, and the business process modeling for production can be seen at Fig. 3.

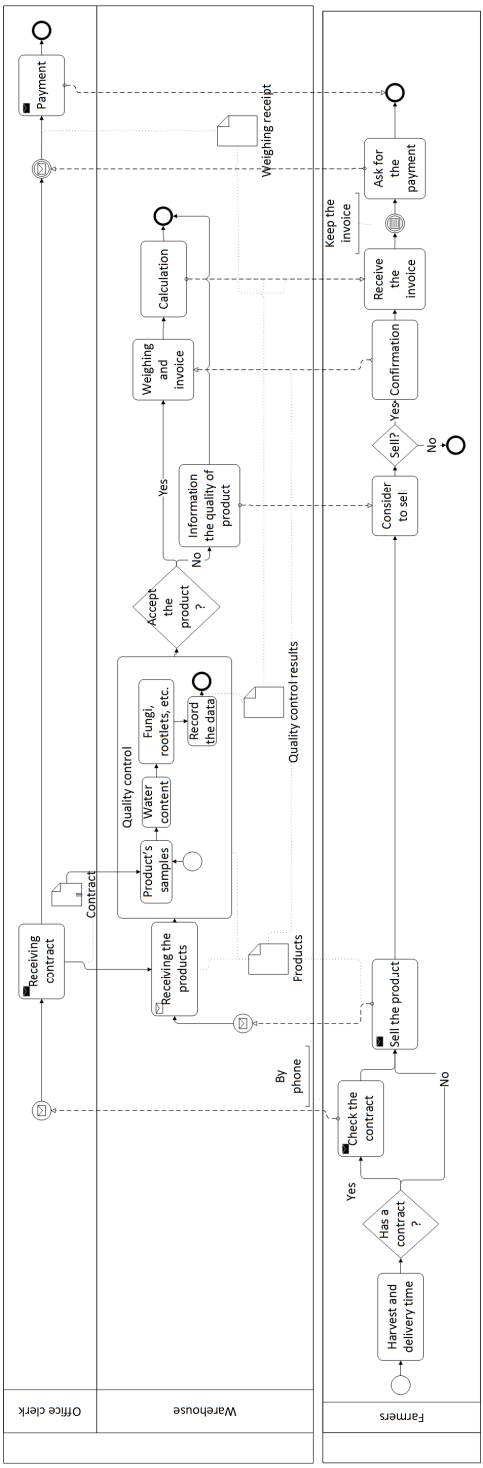


Fig. 2. Business process modelling for purchasing

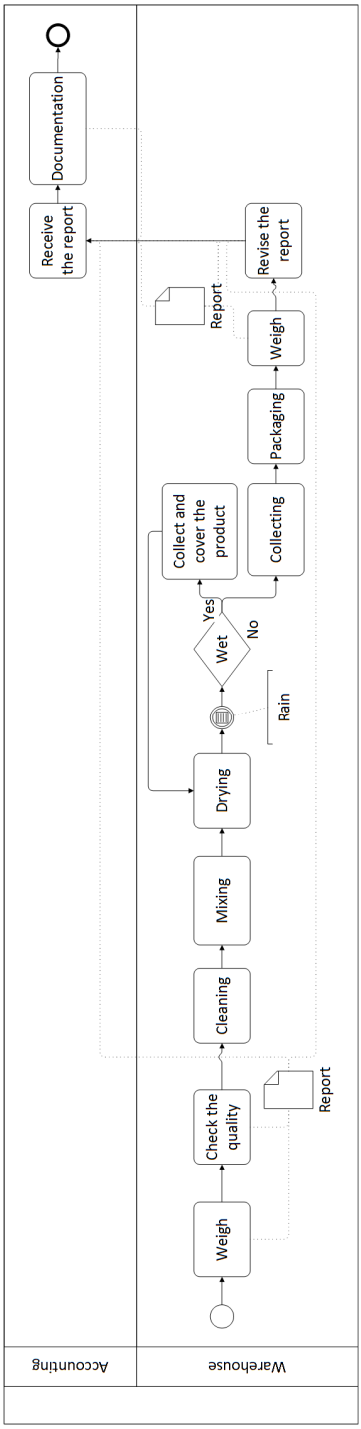


Fig. 3. Business process modeling for production

4.2 Maturity Phases

This research evaluates the maturity phases through six critical success factors which are strategic alignment, culture and leadership, people, governance, methods and informational technology [12].

Strategic Alignment. Strategic alignment is the linkages that connect organizational priorities and enterprise processes in order to achieve its business goals [11, 12]. In this company, the strategy to achieve more gain in profit and operational process already aligned. Based on the data, this company's strategic alignment reaches phase 2.

Culture and Leadership. Company culture implies the usual way that people behave inside. Culture is the values, beliefs, attitudes and behaviors that presents the company identification and public images. Leadership is crucial roles in supporting process design and facilitating cultural change throughout the company [11, 12].

In this agricultural company few years ago, when the management team introduced new business process or changing in operational company, there was certain resistance from some employees, especially employees is used to traditional way to work. The management began to arrange some trainings for changing the way of work. Nowadays, in company, the culture has been established gradually. Employees can accept new business process because they already realized that these changings made their work easier, more effective and efficient. These condition reflects to the phase 2 in Gartner maturity.

People. People are the individuals and groups who continually enhance and apply their knowledge and professional skill related company's business [11, 12]. In this company, staffs know and apply their skill in daily work. Unfortunately, they did not improve their skills, but just did the daily routine works. In same division, staffs could work together, but when they were asked to work cross division, they did not know about the regulation and business process. Based on this analysis, this company stays in phase 1.

Governance. Governance refers to the transparent accountability, responsibility, decision making and reward processes to guide actions [11, 12]. Based on the interview, it proved this company concerned to the regulation in arranging the production and purchasing rules. This company also deal with differences and contradiction between process and policy. As a result, the maturity level of the factor governance arrives phase 2.

Methods. Methods are the approaches and techniques to support consistent process for success of company [11, 12]. When this company wanted to implement a new business process, it will establish a person in charge of the control and improvement. Based on this fact, this company points out that on phase 2.

Information Technology. This factors include the software, hardware and information management systems that enable and support process activities [11, 12]. IT plays

many important roles in this company. Start from the contract with the farmer, finance, warehouse and inventory, this company use IT support to help dealing with data. By the support of IT, this company tries to streamline the process and information flow, which means this company stays on phase 3.

Based on the research findings, finally, it has evaluated the maturity level of the six critical success factors that influence the company. As the Table 2 presents the maturity level of company.

Table 2. BPM maturity level

	Phase 0	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
Strategic alignment	■	■	■			
Culture and leadership	■	■	■			
People	■	■				
Governance	■	■	■			
Methods	■	■	■			
Informational technology	■	■	■	■		

5 Conclusion

Business process modelling can describe the flow of business in this company. This model can help stakeholder conceive the tasks which are going clearly. With the model, stakeholder can make significant improvement in the future. The main problem facing in this agricultural company are: the performance of production process. Based on our analysis, we conclude several advices to help this company better

As the Table 2 provides the maturity level of the people is on phase 1, the maturity level of information technology achieves phase 3, and the maturity level of the other four factors including strategic alignment, culture and leadership, governance and methods have already arrived phase 2. These finding will help company to do big planning for improvement in the future.

In terms of IT, this research suggests company keep IT staying in the phase3 until the other factors reach phase 3 as well. In this way, company can focus more time and resources on the improvement of the other critical success factors.

References

1. Siha, S.M., Saad, G.H.: Business Process Improvement: Empirical Assessment and Extensions. Business Process Management Journal 14(6), 778–802 (2008)
2. Garimella, K., Less, M., Williams, B.: BPM Basics for Dummies, Software AG Special Edition. Wiley Publishing, Inc., Indianapolis (2008)
3. Davenport, T.H., Short, J.E.: The New Industrial Engineering: Information Technology and Business Process Redesign. MIT Sloan Management Review, 11–27 (1990)
4. Van der Aalst, W.M.P., Weske, M., Grunbaeur, D.: Case Handling: A New Paradigm for Business Process Support. Data and Knowledge Engineering 53(2), 129–162 (2005)

5. Ravesteyn, P.: Business Process Management Systems: Hype or New Paradigm. In: Proceedings of the 18th Conference of the International Information Management Association (IIMA 2007), Beijing (2007)
6. Antonucci, Y.L., Goeke, R.J.: Identification of appropriate responsibilities and positions for business process management success: Seeking a valid and reliable framework. Business Process Management Journal 17(1), 127–146 (2011)
7. Van der Aalst, W.M.P., Van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., Weijters, A.J.M.M.: Workflow Mining: A survey of Issues and Approaches. Data and Knowledge Engineering 47(2), 237–267 (2003)
8. Harmon, P.: Business Process Change: A Manager's Guide to Improving, Redesigning, and Automating Processes. Morgan Kaufmann, San Francisco (2003)
9. Allweyer, T.: BPMN 2.0. Introduction to the Standard for Business Process Modeling. Norderstedt (2010)
10. Object Management Group: Business Process Model and Notation (BPMN) Version 2.0. (2011)
11. Gartner: BPM Maturity Model Identifies Six Phases for Successful BPM Adoption. Gartner, Inc. (2006)
12. Melenovsky, M., Sinur, J.: Having a BPM Maturity Model is Important for Long Lasting BPM Success. Business Rules Journal 7(12) (2006)
13. Sugiyono.: Metode penelitian kuantitatif kualitatif dan R&D. Alfabeta, Bandung (2011)

Innovation Strategy Services Delivery: An Empirical Case Study of Academic Information Systems in Higher Education Institution

John Tampil Purba[1] and Rorim Panday[2]

[1] Pengembangan Bisnis dan Manajemen-School of Economics, Jakarta
jpoerba88@gmail.com
[2] Bhayangkara University, Jakarta
indripan@gmail.com

Abstract. Information Communication and Technology (ICT) gives a lot of contributions in the rapid transformation in global society. The people have to adapt in evolving claims of this trend in order to survive from daily struggles. In this 21st century Indonesians compel themselves to go with this flow and have to keep at par with countries over the world. Hence, higher education institutions must undertake on this fast transformation in serving the stakeholders real time by supporting of ICT as at University ABC in Jakarta, Indonesia. The existing Academic Information Systems (AIS) is not yet fully used for moduls usage and some of them are still idle.This paper is an empirical case study and the method used is discriptive research by using the existing data recorded in the systems then analyzed accordingly. This results study will give sugestions for better service delivery to the institutions.

Keywords: innovation strategy, service delivery, AIS, higher education, case study.

1 Introduction

Service delivery becomes the main key strategy in achieving the goals in this era of century development. Most of the organizations over the world always do reviews and redesign on their services delivery toward their customers for the sake of growing and sustainability of their businesses, companies, governments, non-governments organizations and also higher education institutions. For the higher education institutions for example, the presence of information communication technology can support the provision of delivering services with real time output to the need of stakeholders such as transcript of records, load of teaching, syllabus, grades, student finance reports and others immediately. The usage of information communication technology in the institution can be used as a competitive advantage of the institution because of good service delivery to serve their stakeholders. Therefore, the director general of higher education, Ministry of Education Department Republic Indonesia advocated the use of information communication technology in higher education management and

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 514–525, 2015.
DOI: 10.1007/978-3-662-46742-8_47

governance in conducting service management for the quality of services. In this global era, quality administrative is required, as Purba, John Tampil [1], towards the global competitions as well the needs of market place, each higher institution shall evaluate their existing performance of services to face hyper and global competitions. Besides, the regulations from government are also to force any higher education institution to comply all regulations accordingly. The service management shall have quality of all the administrative activities such as on line registration systems even for the new students and also the existing students, subjects, schedule classrooms usage information and teaching learning activities such as; absenteeism, grades, syllabus, lecturers absenteeism, etc.

The AIS data are growing from time to time because of the increasing of the number stakeholders or customers (students) are forcing the institution to improve and upgrade the service delivery in enhancing better access of the stakeholders or customers' needs. This way is chosen to ensure the availability access in delivering the services to all the stakeholders. In providing the regulations, the management shall make such services to meet the quality standard. The people who involved for the service delivery must understand how far the utilization, availability and coverage access to database that are often used. As the purpose of delivering services to meet the requirements of the standards to the stake holders. Anwar, Chairil and Warnars, Spits [2] in serving with the implementation of a service system of learning the academic system is needed and it is a very vital things to implement the learning process continuity and quality. Without an academic information system, the learning process will not running well on learning process. In the world of education is equally important as an example of the card study plan is one component of the terms in the study plan of the students.

Swardi and Permatasari [3], information systems today has become important part of the institution of higher education. Many higher education institutions using information system as a solution of efficiency and competitiveness, but less of them have the ability to accommodate many changes according to the growth of the institution itself. That opinion added by Purba, John Tampil [4] as cited from the statement of Lawson [5], during this period, competitive advantage rested variously on mainstream variables like efficiency, quality, customer responsiveness and speed. In the new millennium, control over the above variables represents the minimum threshold to "play the game". Each factor remains important, but is unlikely of itself or as part of a group to provide a sustainable competitive advantage. Today's organizations face an additional challenge the requirement to innovate, not just occasionally but often, quickly and with a solid success rate. Wohlstter, P., et al [6] the nature of services and applications provided by the information systems group is also believed to have change over time. The new strategic applications also tend to be targeted at the customer or the distribution channel rather than on systems for internal efficiency.

In the case of University ABC Jakarta, one of the big and old universities in Indonesia, in the year 2008 this institution still used semi computerized Academic Information Systems (AIS) with Microsoft Foxpro technology and desktop based application. With such conditions, would certainly be very difficult for the university to be able to compete with other universities that already have a fully computerized of academic information system.

At that time, there were a lot of works such as registration, classroom division, scheduling of lecture, absenteeism and etc, done manually and of course there were a lot reports for internal and external need, became delay. All the students should go to campus to see the grades and other information on provided bulletin boards, and the announcements of the grades from the lecturers are often delayed because of the process of grading is still done manually. In such cases, the demands of the progress of time the presence of technology and information systems needed by an organization that services to stakeholders can be implemented properly. There were a lot of complaints from the students and other stakeholders in regarding the delay of the announcement of their grades and consequences to the delay of re-registration for next batch semester. Noting that the management should make a breakthrough, the new top management of university initiated project on AIS applications in order to provide real-time service delivery to its stakeholders. For that it is needed innovation strategy in building AIS so that can manage all academic information and result optimization the service delivery of information in providing accurate, fast and better services to students in particular and all the stakeholders in general.

2 Literature Review

OECD[7], Innovation is already an important driver of growth in some countries. Firms in several OECD countries now invest as much in intangible assets, such as research and development (R&D), software, databases and skills, as in physical capital, such as equipment or structures. Much multifactor productivity (MFP) growth is linked to innovation and improvements in efficiency. Preliminary estimates indicate that in Austria, Finland, Sweden, the United Kingdom and the United States, investment in intangible assets and MFP growth together accounted for between two-thirds and three-quarters of labor productivity growth between 1995 and 2006, thereby making innovation the main driver of growth. Differences in MFP also account for much of the gap between advanced and emerging countries. This suggests that innovation is also a key source of future growth for emerging economies. (The OECD Innovation Strategy getting a head start on to morrow). The similar shall be implemented in the higher education institution for better productivity by using service delivery.

Strategy to implement the service delivery it is better to scan environment scanning in preparing and calculating the strengths, weaknesses, opportunities and threats. Let us see the strategy requirements; Purba, John T.[8] strategies must be flexible, requiring you to be sensitive and responsive to market changes. The first overriding imperative about markets is that they change. The second imperative is that you must try to know what is going on out there all the time or you will not know what is changing. The third imperative is that if you do not know what is changing you cannot respond. The fourth imperative is that if you cannot respond to changes your strategies will be unrealistic and will not work.

For the university purpose Phillips, Tim [9], Information Technology in all areas: departments, faculties, support services and all other organizational structures and groupings. Therefore all members of the University are stakeholders – students, staff and all others associated with the University of Bristol. IT Services-help, support, training, systems development, business and systems analysis, programme and project

management. All of these require clear, well defined, agreed IT management and governance structures in order to deliver the aims and objectives of this IT Strategy. Implementation of the IT Strategy will be managed through a set of inter-related programmes and projects. The IT Strategy builds on long-standing common sense principles of the Information Strategy. Strategy of the Service delivery will be based on good governance, project and service management, coupled with high quality reliable, sustainable IT Services. Good IT is needed to support the whole University and to enable excellence. Fig. 1 below describes the interrelated the university vision that valued which is truly enhanced by the use of Information Technology:

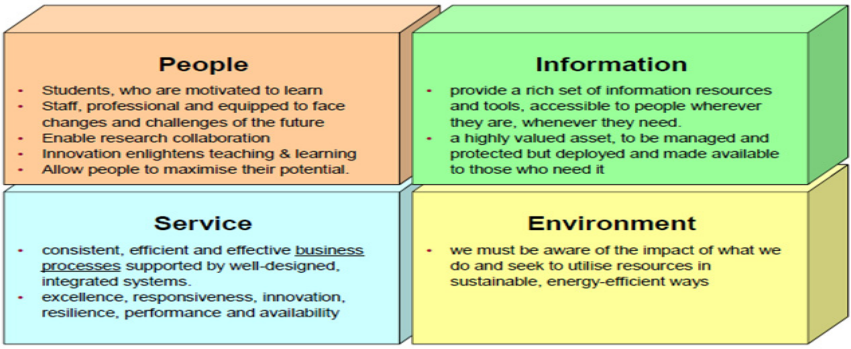


Fig. 1. Vision and Values by the use IT service in University of Bristol [10]

White, Susan [11], Information technology affects not only the delivery of academic content but also ancillary operations. If today's students are more technologically savvy than the majority of the faculty, the implications for the academy surely include a wide range of opportunities to leverage the fruits of new hardware and software tools in ways that truly enhance the learning experience. This will require flexibility on the part of the faculty. In addition, as schools are more ratings conscious, new ways to innovate in the delivery of supporting services will emerge; institutions whose operating processes are most open to change will reap the most benefits. Purba, John Tampil [12], all most modern organizations implement and rely on Information Technology services to support their business processes and sustainability. As we know IT services are built upon the technical infrastructure, systems and application software. The set of processes that apply planning, organizing, directing and controlling the provisioning of IT services. Besides of the ICT development inside the organization the human resources shall also develop in achieving the good service innovation as such opinion as described by Six Sigma [13] at the heart of every services business are the opinions, behaviors and decisions made by people. Also of great importance in services settings are work processes, whether they are recognized or not. Until a process focus–rather than a task focus–is developed, the scope and "sticking power" of improvements will be limited. Analyzing and modifying human performance in these environments is as complex as any manufacturing situation – but the tools and methodology required to achieve the legendary improvements of Six Sigma are significantly different. Six Sigma for Services Companies Training is the first program designed specifically for professional services providers.

Fig. 2. Service innovation for delivery of existing services;Sources: Six Sigma Qualtec 2014

In delivering services to the stakeholders and customers the institution shall provide and concern on managing performance, managing people, managing resources, conduct planning, and managing services. Those are the requirements for working which others, improving services, setting direction, and demonstrating personal qualities. The figure below displays about them.

Fig. 3. Delivering the services, Sources: Green, Matt and Gell, Lynne (2012)

In the second article of Green, Matt and Gell, Lynne [14] comprises four subdomains: planning, managing resources, managing people, and managing performance. Ensuring the correct processes and infrastructure are in place will help to ensure that the services you oversee are fit for purpose. Any business planning process should be informed by patient or service user feedback, audit results, commissioners' intentions, staff input, and best practice evidence. For Managing performance to achieve exceptional results can use performance management techniques to support, energize, and empower staff to improve their performance. A clear understanding of the principles of performance management of individuals should be sought, such as supporting poorly performing colleagues to improve through regular competency based review meetings to monitor progress.

Another opinion on service delivery in the integration with ITIL service management is the combination of service support and service delivery. The service delivery contains; service management, financial mgt of IT services, capacity management, IT service continuity management, and availability management. The service delivery

seems integrated with information technology driven. The integration of all services as display in the below figure.

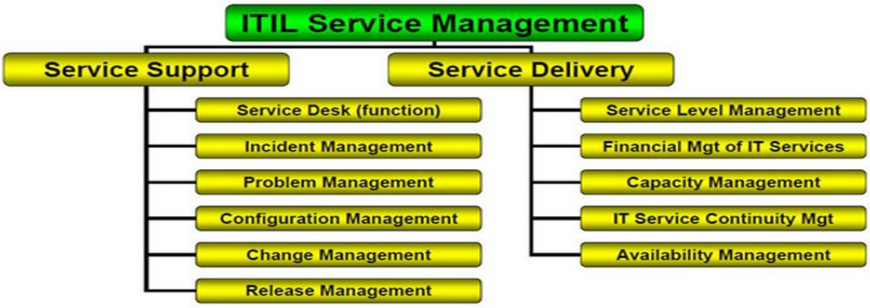


Fig. 4. ITIL Service Management (Cousin, Carlos, 2010)

Cousin, Carlos [15] Information Technology Infrastructure Library (ITIL) is recognized as the de facto standard for IT Service Management. ITIL reflects a process-model based view of controlling and managing IT functions. The approach is credited to W. Edwards Deming and his plan-do-check-act (PDCA) cycle. In 2009, the ITIL library consisted of the following 9 volumes: Service Support, Service Delivery, ICT Infrastructure Management, Security Management, The Business Perspective, Application Management, Software Asset Management, Planning to Implement Service Management & ITIL Small-Scale Implementation (which was added to the original 8 volumes). Three Key Objectives of ITIL are to align IT services to meet the needs of business and customers, Improve quality of IT services delivered and reduce the long-term cost of service. He continued that providing leadership of services and staff, consultants are also expected to pioneer and drive forward new models of service delivery.

3 Research Methodology

To explore the service delivery for Academic Information Systems (AIS) that has been implemented in University ABC Jakarta, Indonesia, the researcher used the existing data which have been recorded in the system that can be accessed accordingly. Through collecting and selecting the data about how the service delivery of Academic Information System that utilized by management as well as lecturers and stakeholders. The methodology used in this paper is descriptively. The descriptive analysis was done by giving the flows of delivering service data to the stakeholders and also by selecting and capturing the data related such as; figures and tables from the existing display data in the applications then analyzed and discussed accordingly.

4 Results and Discussion

As the one of the old big university in Jakarta the new mangement did innovation to overcome the problems raised from the stakeholders, so the Academic Information

System in Web based technology become their priority in the first year of services. Then they set up committees to do the AIS application technology by appointing the head information technology bureau to manage the project.

In business process, herewith the researcher describes flow of delivering services to the stakeholders; students, lecturers/faculty members, parents and management of the university are as the flow of the figure below. All data are stored in the server data storage which have been input by the data entry staffs in the colleges level. The total departments in the university are 25 which aproximately five thousands students. They are served whith eye contact teaching and learning activities, either in classrooms or in the laboratoriums.

Fig. 5 below is flow of information and communication technology of the Academic Information System in University ABC Jakarta. The services continue management systems are prepared in a big data with cloudly web based technology that can be retrieved in seven (7) days and twenty four (24) hours. The technology choiced due to the strategic policy of the rector of the university at that time.

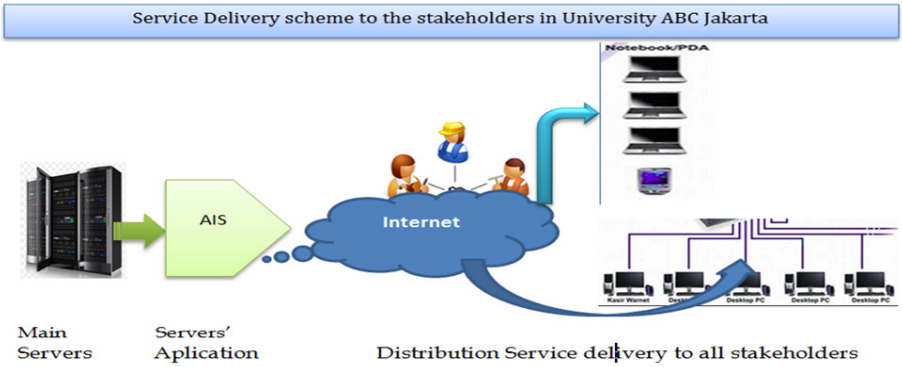


Fig. 5. AIS Service Delivery scheme

In the teaching and learning activities are all recorded and storages in the server, such as the absenteism of students as well as lecturers from time to time and ofcourse automatically displayed in screen of computers, PDAs, notebooks, IP phones and other modern devices that suitable to the systems. The services of AIS were developed with Microsoft SQL Server 2008 data base application for the back end, and in front end with programming language Microsoft C# dot net.The persons of data entry staffs who responsible in every college they are also stay in the office of colleges to input data of the daily absenteisms of the students and lecturers. For grades of assignments, mid semester, and final semester shall be input into the system by the lecturers according their subjects teaching.The students can continue to communicate with lecturers anytime, anywhere by accessing to the AIS server system. Such systems are not only to speed up services delivery but also would help ease the burden of staff in the teaching-learning process. This can be trusted because in this system some staffs functions could be taken over in a computer programmes which known as agent sytems.

In delivering services to the stake holders the management of the institution provides the good and high end ICT infrastructures such as hardware, equipments, antennas, and others devices which connected to the systems database. For availability management and service level management, the infrastructure communication inside and outside the campus supported by outside vendors such as : Internet Provider company, equipment suppliers, software companies and others. All of these combined with collaboration integration accordingly in giving service delivery to the stakeholders. The schematic of the infrastructures designed by innovative thinking with the existing situation and condition with more the one campus sites. The below figure can demonstrate existing ICT communication data infrastructures.

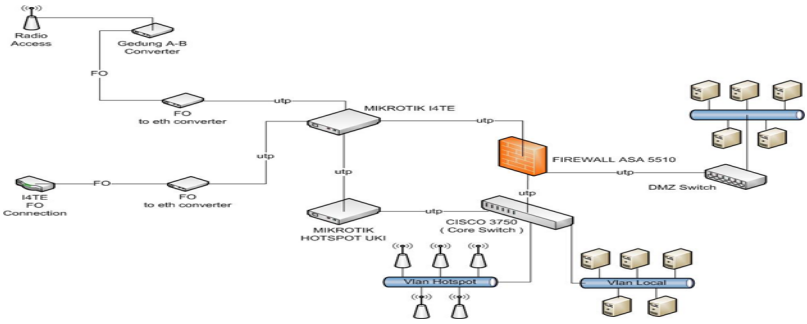


Fig. 6. Internet Provider service delivery inside the campus (University ABC Jakarta)

Fig. 6 above illustrates how the service delivery to the stakeholders by using the new ICT infrastuctures which availabel in the site of the both campus in order to accomodate the needs of them. The network infrastucture has been used and managed day to day by the department of network and infrastures of IT Division according to standard operating systems. There is a service desk avail in the department to handle if there is problems arised then continued to related IT engineer for solving the problems accordingly.

Fig. 7 below displays the modules of Academic, Students, Parents, Faculty members/lecturers, and for new enrollment students. The academic one give many services on line; this module used by top management, administrators, deans, head of bureaus, and others as specified in the regulation of hierarchies. This module can be used to give information which categorized public information, so the students and all stakeholders can use to see and to know the latest news in the campus, the example as displayed below:

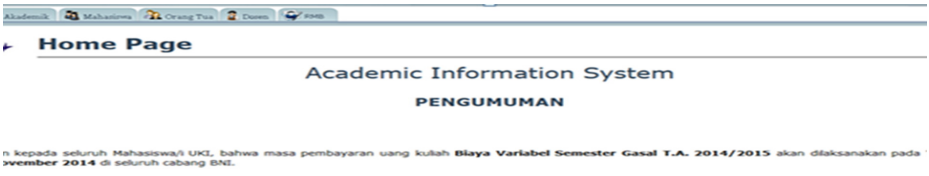


Fig. 7. Portal Academic Information Systems of University ABC JakartaSource: Portal University ABC, Jakarta Indonesia [16]

The module of students provides to serve to students which giving the information to such as; grades, finance, on line registration, syllabus, home assignments, study planning and others. In the module of parents provides the information about their children absenteeism, grades, finance (including the payments history) and others related. For lecturers/faculty members modules are displayed in Fig. 8 below; the lecturers can change the profiles such as password, address, researches, community development, seminars, syllabus, and others. This module also can be used to see the subject teaching loads records, absenteeism records. Besides the lecturers can used this module in giving information, home assignments, verification for students on line registrations, students' leaves and entry the students' grades. In addition, the results of Study Plan Card (Kartu Rencana Studi) is accountable to any party or parties both students of the university because of the process in accordance with the rules that have been standardized and enforced by the management. The stored data in the servers as set with the formula in database, can be used by the students and academic agencies such as the management of the institution itself. With the presence of these web-based AIS technology constraints contained in the system is done manually can be resolved. It is an idea to create software applications that are able to manage the administration of institution well, so that these constraints can be handled with this software. Having this software application is expected to provide solutions problems of service delivery quickly and accurately such as the Study Plan Card (KRS). So the expectation of services based on information systems can be implemented, especially to serve the students as customers must be served better and satisfactorily.

Fig. 8. Module of Lecturer in the Academic Information Systems Portal ABC University

The AIS application can be used by the lecturers to give the appropriate service delivery to the students, and for their Beban Kerja Dosen's report (lecturer's working load report) that give to the government each semesteral period. Each lecturer can optimize this application systems many purposes as regulated by the inside institution and Department of Education and Culture or Department of Research and Higher Education Republic Indonesia as the regulator for education in the country.

Table 1 below displays the present or absenteeism of the lecturers and the students during the teaching and learning process in one semester. This system not available before the year of 2009 in the institution, all works like these were done manually so they took many hours to calculate the number of days or class attendance of the

students in one semester. As per regulation while the students cannot fulfill 75 percent of all classrooms attendance, the student(s) prohibited to take the final exam. As display in the below table the percentage of the attendance were automatically calculate by the machine of the AIS systems. This new systems is very helpful as the service level management for all respective management to know the accountabilities of all the lecturers and students performance especially from the classrooms attendance.

Table 1. Display of the attendances of lecturers and students

07:10 - 09:40

	Pertemuan1 9/13/2011	Pertemuan2 9/20/2011	Pertemuan3 9/27/2011	Pertemuan4 10/11/2011	Pertemuan5 10/18/2011	Pertemuan6 10/27/2011	Pertemuan7 10/28/2011	Pertemuan8 11/22/2011	Pertemuan9 11/29/2011	Pertemuan10 12/06/2011	Pertemuan11 12/13/2011	Pertemuan12 01/03/2012	Pertemuan13 01/10/2012	Pertemuan14 1/13/2012	Hadir	Alpha	Sakit	Persentasi Kehadiran
931150040	Alpha	Alpha	Alpha	Hadir	Alpha	Alpha	Hadir	Alpha	Alpha	Alpha	Alpha	Alpha	Hadir	Hadir	4	10	0	26.66%
1031150034	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Alpha	Hadir	Hadir	Hadir	Hadir	Hadir	13	1	0	86.66%
1031150030	Alpha	Alpha	Alpha	Hadir	Alpha	Alpha	Hadir	Alpha	Alpha	Alpha	Alpha	Alpha	Alpha	Alpha	2	12	0	13.33%
1131150034	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Alpha	Hadir	Hadir	Alpha	Alpha	Hadir	Hadir	11	3	0	73.33%
1131150056	Alpha	Alpha	Hadir	Alpha	Alpha	Alpha	Hadir	Alpha	Alpha	Alpha	Alpha	Alpha	Alpha	Alpha	2	12	0	13.33%
1131150801	Hadir	Hadir	Hadir	Alpha	Sakit	Sakit	Hadir	Alpha	Sakit	Sakit	Alpha	Alpha	Sakit	Sakit	4	4	6	66.66%
1031150014	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	14	0	0	93.33%
1131150016	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	14	0	0	93.33%
1131150035	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Alpha	Hadir	Hadir	13	1	0	86.66%
1131150042	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Alpha	Hadir	Hadir	Hadir	Hadir	Hadir	13	1	0	86.66%
1131150068	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Sakit	Hadir	Hadir	Alpha	Hadir	12	1	1	86.66%
1131150029	Hadir	Hadir	Hadir	Alpha	Hadir	Hadir	Hadir	Hadir	Alpha	Hadir	Hadir	Alpha	Hadir	Hadir	11	3	0	73.33%
1131150010	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	14	0	0	93.33%
1131150042	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Alpha	Alpha	Hadir	Hadir	12	2	0	80%
1131150092	Hadir	Alpha	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Alpha	Hadir	Sakit	Hadir	Hadir	11	2	1	80%
1131150077	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	14	0	0	93.33%
1031150070	Alpha	Alpha	Hadir	Hadir	Hadir	Hadir	Hadir	Alpha	Hadir	Hadir	Hadir	Hadir	Hadir	Hadir	11	3	0	73.33%
1131150078	Alpha	Alpha	Hadir	Hadir	Hadir	Hadir	Hadir	Alpha	Hadir	Hadir	Alpha	Hadir	Hadir	Hadir	10	4	0	66.66%

Table 2. Information of grades entried by lecturer in the AIS module, ABC University Jakarta

rang Tua | Dosen | RMB

MEMASUKKAN NILAI

Tahun Akademik	2010/2011 - Tahun Akademik 2010/2011
Semester	Gasal
Kelas Mata Kuliah	31219 - Sistim Informasi Manajemen - A - Jumat - 07:10
Proses	Proses

No	NIM	Nama	Nilai Tugas	Nilai UTS	Nilai UAS	Nilai Akhir	Nilai Huruf
1	0431150013	Dicky Roy Mangara	69.50	63.00	63.00	64.30	B-
2	0531150041	Samuel Andra	67.20	50.00	65.00	60.94	B-
3	0631150020	Nicholas Sahala Pohan	54.70	50.00	60.00	55.94	C+
4	0631150035	Sunarni Alimbalu	0.00	0.00	0.00	0.00	0
5	0631150065	M.Rizal	58.00	48.00	68.00	60.00	B-
6	0631150068	Tessa	48.00	63.00	70.00	63.50	B-
7	0831150002	Abner Esau Runpah Ataupah	75.00	88.00	80.00	81.40	A
8	0831150004	Enyda Debora Sitio	73.40	62.00	78.00	72.28	B+
9	0831150005	Monika Andriana Sitorus	75.50	78.00	73.00	75.00	A-
10	0831150009	Parinah	70.30	75.00	78.50	75.81	A-
11	0831150012	Togi Mangapul Sitinjak	71.80	63.00	70.00	68.26	B
12	0831150014	Irwandi	0.00	20.00	0.00	0.00	0
13	0831150018	Lieliyana Octavia	78.00	75.00	72.00	74.10	B+
14	0831150019	Kristian Daviyanto	69.00	62.00	72.00	68.40	B
15	0831150020	Elvania Margareth Tanamal	0.00	68.00	0.00	0.00	0
16	0831150021	Andry Chrisnanto	73.60	83.50	85.00	82.27	A
17	0831150022	Eunike Priskila	75.00	62.00	73.30	70.25	B+
18	0831150023	Anton Pratama	67.00	45.00	58.00	55.90	C+
19	0831150024	Hercia Resita	67.50	64.00	60.00	62.70	B-

As displayed in Table 2 above, we see how the systems can deliver the services for the students' grades, when the lecturers do entry data of grades and submit in the systems directly the AIS application display them and each student can see his/her grades real time anywhere and everywhere as far as devices and internet connection available. There are many modules were available in the systems but there many of them still idle, so they need more and more trainings in order to optimize the systems. This system is a new ones which new technology that there is not available yet before

the year 2009 with the new and high technologies. The services are very helpful ones for all the students and management as well. In accordance with the idle module the management of the institution can hire or add staffs which have been familiar with this system in optimizing the idle ones. Then the management, stakeholders and also board trustees can use this application system to control and monitoring the activities especially in academic perspectives.

5 Conclusion

Recently regulations of the higher education institutions forces to do the good university governance. The reports of any university shall give periodically to the government and evaluated by the government officers from time to time. The evaluation has consequences to rank of the institutions and operationally licences for the future. This paper discusses about the usage of ICT in conducting the service delivery in higher education institution which used university ABC Jakarta as an empirical case study that can be used in other higher institution over all the country.

The service delivery by using information technology with high end standard is highly recommended in giving the real time service delivery to the stakeholders. The management of the institution suggests implementing the innovation strategy by using AIS application technology in providing the service to the students and lectures also other stakeholders. The results of this study in the paper, the researchers propose and recommend university ABC to optimize the idle module and sub-module in application in giving better service delivery to the stake holders. There are a lot of higher education institutions in the wide country still not yet use the dot net or web based technology application, that consequently making the unhappy students and late reports to the government, the other higher education institutions can adopt this system to manage the services delivery in their higher institutions.

References

1. Purba, J.T.: Service Management Strategy by Implementing AIS in Higher Education in Indonesia: Case Study Christian University of Indonesia. In: Proceedings of the 11th UBAYA International Annual Symphosium on Management (2014)
2. Anwar, C., Warnars, S.: Sistim Informasi Akademik Online Sebagai Penunjang Sistem Perkuliahan (2008), http://arxiv.org/ftp/arxiv/papers/1006/1006.2082.pdf (accessed November 1, 2014)
3. Suwardi, I., Permatasari, D.S.: New Integration Model of Information System on Higher Education. In: Proceedings of the International Conference on Electrical Engineering and Informatics Institut Teknologi Bandung, Indonesia, June 17-19 (2007)
4. Purba, J.T.: PengukuranKualitas Layanan Unit KerjaTeknologi Informasi Menuju Strategi Tata Kelola Organisasi: Studi Kasus. Paper presented, in the Seminar Nasional dan Call paper, Fakultas Ekonomi Universitas Maranatha, Bandung Indonesia (2014)
5. Lawson, B., Samson, D.: Developing Innovation Capability in Organization: A Dynamic Capabilities Approach. International Journal of Innovation Management 5(3), 377–400 (2001)

6. Wohlstter, P., et al.: Improving service delivery in Education through collaboration: An Exploratory study of the role of Cross-Sectoral Alliances in the Developement and Support of Charter Schools (2004)
7. OECD.: Service DeliveryIn Fragile Situations, Key Concepts, Finding and Lessons (2010), http://www.oecd.org/development/incaf/40886707.pdf (accessed October 2, 2014)
8. Purba, J.T.: Proposed Strategic Plan for Faculty Development for Atmajaya Catholic University Jakarta. The Doctoral Dissertation. De La Salle University, Taft Avanue, Vitocruz Manila Phillipines, pp. 11–12 (2002)
9. Phillips, T.: Information Technology Strategy Supporting Excellence, University of Bristol.Manual book. Version 3.1 (2008)
10. Purba, J.T.: Usage of IT Servicesin Higher Education Management for Innovation Strategy: A Case Study. Paper presented in the International Coference of Organizational Innovation (ICOI), De La Salle University, Manila, Philippines, August 12-14 (2014)
11. White, S.C., Glickman, T.S.: Directions for Higher Educations, no. 137, Spring ©Wiley Periodicals, Inc. Published online A in Wiley Inter Science (2007), http://www.interscience.wiley.com, doi:10.1002/he.248 (accessed October 10, 2014)
12. Purba, J.T.: Service Performance ICT Train New Technologies Towards Orgazizations: Case Study Asindo Informatika. Paper presented in the Seminar Nasional dan Call Paper "Business in Society: Towards Asian Era", Fakultas Ekonomi dan Bisnis Universitas Sebelas Maret, Surakarta, March 18-19 (2014)
13. Six Sigma.: Business Performance and Process Improvement Solutions (2012), http://www.ssqi.com/solutions/scorecards.html (accessed October 25, 2014)
14. Green, M., Gell, L.: Effective medical leadership for consultants: managing services, improving services, and setting direction. Second article (November 2012), http://careers.bmj.com/careers/advice/view-article.html?id=20009862 (accessed September 29, 2014)
15. Cousin, C.: ITIL Service Management can be viewed as two distinct management sets from the Service Delivery and Service Support (2010), http://carloscousin.com/wordpress/?p=283 (accessed October 27, 2014)
16. Cousin, C.: Portal Academic Information System, University ABC, Jakarta, Indonesia and extrareport (2014), https://ais.abc.ac.id (accessed November 21, 2014)

Intelligent Applications

Public Transport Information System Using Android

Agustinus Noertjahyana, Gregorius Satia Budhi,
and Agustinus Darmawan Andilolo

Informatics Engineering, Petra Christian University
Jl. Siwalankerto 121-131 Surabaya, 60236
{agust,greg}@petra.ac.id, m26409133@john.petra.ac.id

Abstract. Traffic jams are getting higher, making people surabaya thinking of switching from private to public transport vehicles. But the problem that arises is the lack of information about public transportation in Surabaya, so it is quite difficult for people who want to use public transport for transportation.

With the development of technology, especially in the smartphone, which is almost used by most people, the idea emerged to develop an application that can provide information related to public transport service based on mobile technology. Android platform chosen for the current smartphones starting from the lowest price to the highest price is dominated by the Android operating system. Applications developed to help people to be able to choose the transport used while traveling from one place to a particular destination with as little effort as possible. The application is able to assist the selection of an appropriate route, either directly or indirectly.

Keywords: Android, Information System Transportation, Mobile Device.

1 Background

Android-based smart phone technology is increasingly being used by the community. Various types and brands of smart phones that both local and global enliven the smart phone market.

The development of technology and prices are getting cheaper, smart phone with cheap price also is equipped with GPS (Global Positioning System) [1], a technology that can help to find a location from anywhere. This is very helpful in providing information about a location.

Surabaya as the second largest city in Indonesia, have problems in the public transport system. Many people began to leave the public transport as the main option in meeting the transportation needs because quite difficult to get information about the route of public transport [2].

So many public transportation but very little information is available either through the mass media and the Internet. For those reasons, by utilizing the Android technology, where this technology is quite cheap and more people are using it, then this is a pretty interesting idea to be able to develop applications that can help to provide information about public transport service in the city of Surabaya [3].

If the system is developed successfully, it will be easily adopted to be implemented in other areas, so as to help the government to further improve public transport facilities for the people of Indonesia in general.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 529–538, 2015.
DOI: 10.1007/978-3-662-46742-8_48

Applications developed will be able to help people in Surabaya to maximally utilize public transportation, and even not only the indigenous people of Surabaya but also newcomers can take advantage of public transportation in Surabaya [3].

2 Theory

2.1 Public Transport

Transportation system plays an important role in economic development in an area. The better the transportation system of a region, it will have a positive impact on the economic development of the region [4].

Surabaya, the capital of East Java province, is the second largest city in Indonesia after Jakarta. Surabaya city has grown rapidly in terms of trade and population growth. This leads to the need of transportation have increased significantly. The number of private vehicles, according to data of East Java Police has reached more than 4 million units [5].

With the number of private vehicles is large enough, then Surabaya experiencing traffic congestion problems. To overcome traffic congestion, then the people of Surabaya should be encouraged to switch to public transport. For those reasons, the level of comfort and reliability of the public transport system of choice for workers, students and the community at large.

The main problem in Surabaya is information about public transport. People in Surabaya have difficulty in accessing information about public transportation route, so there is a tendency to use private vehicles.

2.2 Public Transport Profile

Public Transportation in Surabaya consist of: City Bus, Public Passenger Cars (MPU), trains and ships. However, the focus of this research is on Passenger Cars (MPU). Each route has a specific code that is set by Organda (Organisation of National Transport Vehicle On Road). List of routes MPU in Surabaya is in Table 1 [6].

Table 1.

Code	Route
C	Pasar Loak - Sedayu - Karang Menjangan PP
D	Joyoboyo - Pasar Turi - Sidorame PP
E	Petojo - Tanjungsari - Balongsari PP
F	Joyoboyo - Pegirian - Endrosono PP
G	Joyoboyo - Karang Menjangan / Lakarsantri / Karang Pilang PP
H2	Pasar Wonokromo - Pagesangan PP
H2P	Pasar Wonokromo - Terminal Menanggal PP
I	Kupang - Benowo PP
K	Ujung Baru - Kalimas Barat / Pasar Loak PP
L2	Ujung Baru - Sasak - Petojo PP
M	Joyoboyo - Dinoyo - Kayun - Kalimas Barat PP
N	Kalimas Barat - Menur - Bratang PP

Table 1. (*Continued*)

O	Tambak Wedi - Petojo - Keputih PP
O1	Kalimas Barat - Keputih PP
O2 (WK)	Tambak Oso Wilangun (Depan SPBU) - Petojo PP / Tambak Wedi - Keputih - Bratang PP
P	Joyoboyo - Gebang Putih - Kenjeran / Petojo - Ketintang / Joyoboyo - Karang Menjangan - Kenjeran PP
Q	Kalimas Barat - Bratang PP
R	Kalimas Barat - Kapasan - Kenjeran PP
S	Joyoboyo - Bratang - Kenjeran PP
T1	Margorejo - Joyoboyo - Sawahan - Pasar Loak PP
T2	Joyoboyo - Mulyosari - Kenjeran PP
U	Joyoboyo - Rungkut - Wonorejo / Joyobekti PP
V	Joyoboyo - Tambakrejo PP
W	Dukuh Kupang - Kapas Krampung - Kenjeran PP
Y	Joyoboyo - Demak PP
Z	Kalimas Barat - Benowo PP
TV	Joyoboyo - Citra Raya / Manukan Kulon / Banjar Sugihan
DP	Kalimas Barat / Petekan - Manukan Kulon PP
Z1	Benowo - Ujung Baru PP
J	Joyoboyo - Kalianak PP
BK	Bangkingan - Karang Pilang PP
DA	Kalimas Barat - Citra Raya
JTK	Joyoboyo - Tambak Klanggri PP
JTK2	Joyoboyo - Medokan Ayu PP
R1	Kalimas Barat - Nambangan - Kenjeran PP
WLD	Wonoarum - Pasar Loak - Dukuh Kupang PP
WLD2	Bulak Banteng - Dukuh Kupang PP
RT	Rungkut - Pasar Turi PP
LMJ	Lakarsantri - Manukan Kulon - Kalimas Barat PP
BM	Bratang - Perumnas Menanggal PP
JBMN	Joyoboyo - Gunung Anyar PP
LK	Manukan Kulon - Pasar Loak - Kenjeran PP
GL	Pasar Loak - Gadung PP
JK	Joyoboyo - Kalijudan - Kenjeran PP
IM	Benowo – Simokerto
WB	Wonosari - Bratang PP
DKM	Dukuh Kupang - Menanggal PP
DKB	Dukuh Kupang - Benowo PP
BJ	Benowo - Kalimas Barat PP
RDK	Dukuh Kupang - Benowo PP
UBB	Ujung Baru - Bratang PP
UBK	Ujung Baru - Kenjeran PP
JMK	Kenjeran - Kalimas Barat PP
KIP1	Kutisari Indah - Petojo PP (Lewat Tengah) PP
KIP2	Kutisari Indah - Petojo PP (Lewat Timur) PP
GS	Gunung Anyar - Sidorame PP
RBK	Rungkut Barata - Kenjeran PP
DWM	Balongsari - Pangkalan Karah PP

2.3 Android

Android is a software for smartphones that includes an operating system, middleware and key applications that are released by Google [7].

To make it easier to develop applications on the Android platform requires Adroid SDK, Tools API (Application Programming Interface) using the Java programming language.

Android Development Tools (ADT) is a plugin that is designed for the Eclipse IDE that make it easier to develop Android applications. By using ADT for Eclipse, it will be easier to create applications Android project, create a GUI application, and add other components. In addition, it can build Android package (.apk) that are used to distribute Android applications.

2.4 Google MAPS

Google Maps is an online map application service provided by Google for free. Google Maps officially accessible through the site http://maps.google.com. Service is very interactive google maps, maps in the slide according to user needs and the level of zoom can be adjusted [8].

Google Maps has many features that can be used eg location search based on certain keywords, such as a name, city, or road, and able to perform the calculation of the travel route from the starting location to the destination.

3 Design System

3.1 System Architecture

Referring to the public transport route as in Table 1, the transport information system developed using google maps to perform routes calculations, and data is saved to the server that can be accessed via the Internet.

There are two functions in the system: administrator and user. The administrator in charge to enter all the data into routes server, while the user is only able to request to the server on a particular routes, and the server provides the results to the user.

System architecture diagram can be seen in Figure 1.

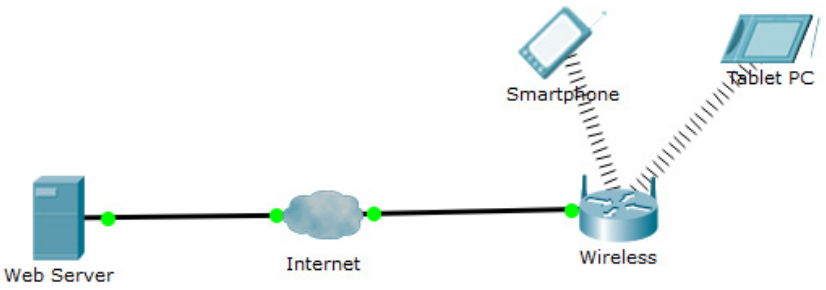


Fig. 1. System Architecture Diagram

3.2 Design Flowchart

Design Flowchart of Administrator Function can be seen in Figure 2.

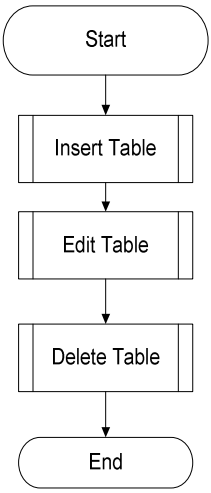


Fig. 2. Flowchart of Administrator Function

3.3 System on Android

Android system begins with the way the user make a selection starting location and destination.

At the destination, the user has three options: choose a destination by street name, place name, or select from the map.

Next the system will process and deliver the results to the user. When user presses the button MAP, then the user will be taken to a page that displays the position of each path traversed by public transport.

4 Implementation

4.1 Web Administrator

On the Web Administrator, there is a page to add street code and name of the street. When Insert button is pressed, the new data will be entered into the database.

At the bottom there is a function to add the location. This function is in the form of ID location that serves to explain the location of a street, street ID based on the code of the street that has been defined at the top.

Latitude and Longitude are used to store data based on the existing location on the map. The function of the web administrator can be seen in Figure 3.

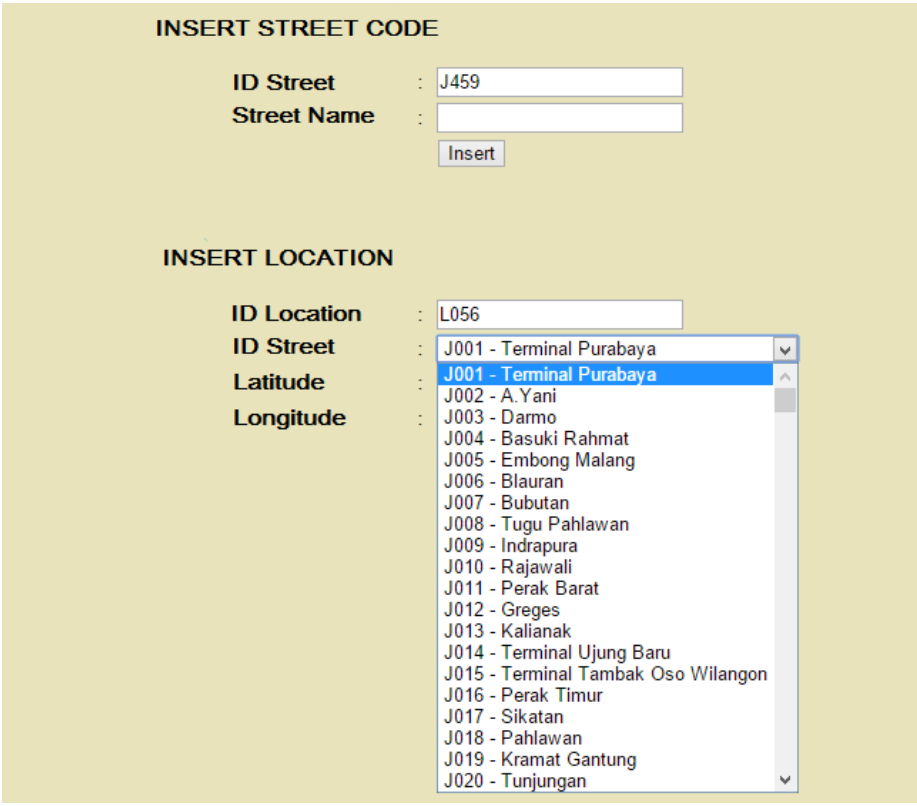


Fig. 3. Web Administrator

4.2 Implementation on Android

Implementation of Transport Information System in the form of Android-based applications. The main page of the application can be seen in Figure 4. There is an input to enter the starting location (Lokasi Awal), or take a position on the map (Ambil Posisi).

For the location of Interest, there are two options, namely by street or by location. If the user selects a street, by entering the name of the destination street. If the user selects a location, by entering a name in the text box. Process of the user to determine the location of the destination by street or location can be seen in Figure 5.

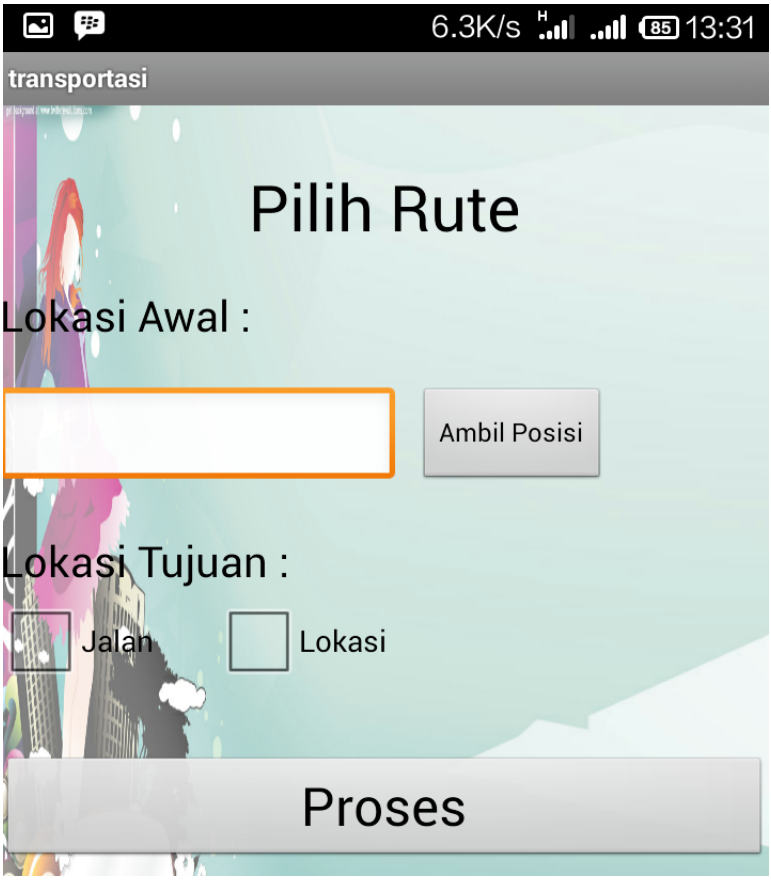


Fig. 4. Main Page of Application

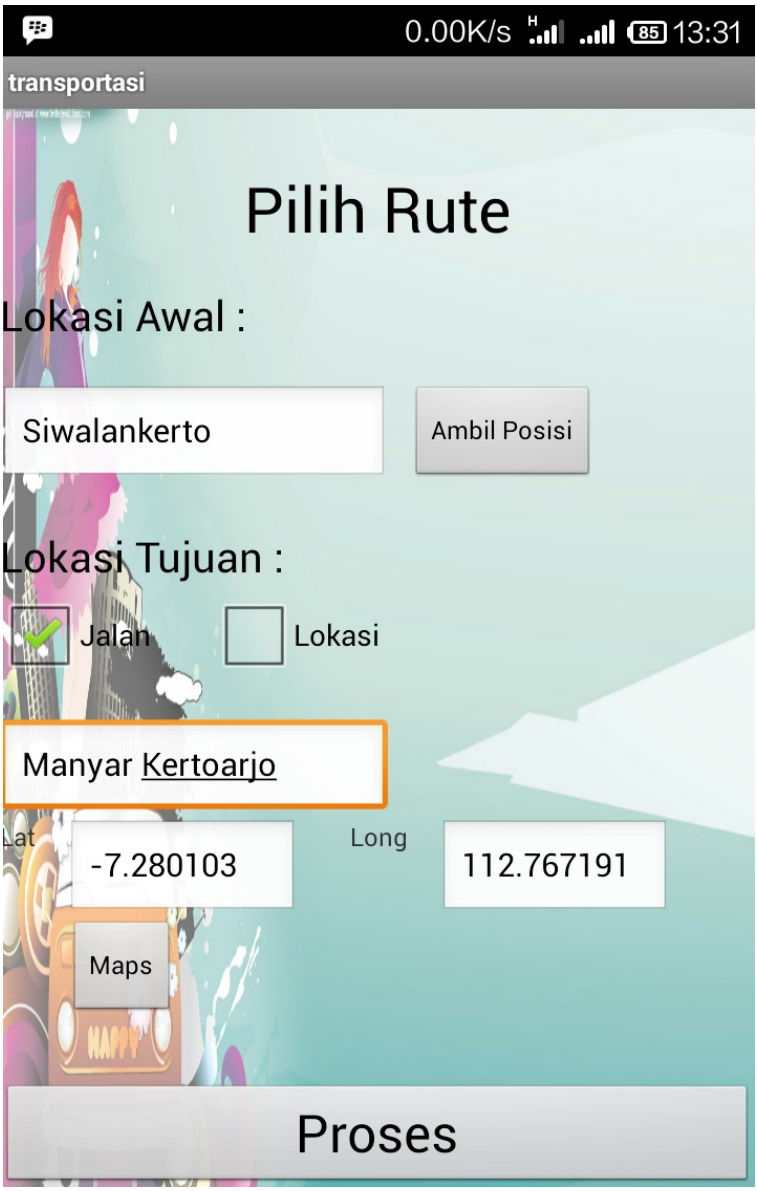


Fig. 5. User selection of Destination

Search results based on the location of the start and destination can be seen in Figure 6.

Fig. 6. Search Result

5 Conclusion

Based on testing performed, the application is able to perform these searches, either directly or indirectly. However, to search for a location, can not be precise because of the limitations of existing public transport. Advice that can be given is the development using a variety of public transportation, not only one kind only.

References

1. Anderson, R.E.: Building a Transportation Information System Using Only GPS and Basic SMS Infrastructure (2009), http://depts.washington.edu/ddi/publications/ictd_2009_starbus_final.pdf
2. Dwiky, A.: Desain Dan Implementasi Sistem Informasi Rute Angkutan Umum Daerah Surabaya Menggunakan Visual Basic 6. Surabaya. UPN Veteran, Jawa Timur (2012)
3. Andilolo, D.A., Agustinus, N., Gregorius, S.: Perancangan Dan Pembuatan Aplikasi Informasi Transportasi Umum Di Surabaya Berbasis Android Surabaya. Universitas Kristen Petra (2014)
4. Weisbrod, G.: Models to predict the economic development impact of transportation projects: historical experience and new applications. The Annals of Regional Science 42(3) (September 2008)
5. Anonim (2013), http://www.surabayakita.com/index.php?option=com_content&view=article&id=5679:tambah-tahun-surabaya-tambah-macet&catid=25&Itemid=0
6. Profil Transportasi Kota Surabaya, http://www.surabaya.go.id/infokota/index.php?id=7 (last accessed December 29, 2014)
7. Safaat, N.: Pemograman Aplikasi *Mobile Smartphone* and *Tablet* PC Berbasis Android. Penerbit Informatika, Bandung (2012)
8. Shodiq, A.: Tutorial Dasar Pemograman Google Maps API (2012)

Lecturers and Students Technology Readiness in implementing Services Delivery of Academic Information System in Higher Education Institution: A Case Study

Rorim Panday[1] and John Tampil Purba [2]

[1]Bhayangkara Jaya University, Management, Jakarta, Indonesia
indripan@gmail.com
[2]STIE-PBM, Management, Jakarta, Indonesia
jpoerba88@gmail.com

Abstract. Now, ICT is a part of human needs in every activity, including education. Academic information systems in Indonesia, has already implementing ICT, either partially or as a totally. How well the information system is created, will depend on the readiness of the stakeholders in Higher Education, especially lecturers and students. This study aims to reveal the Technology Readiness (TR) of Lecturers and students in the academic system implementation. This study refered to the TR that developed by Parasuraman and Colby. This research conducted at the XYZ university, located in Jakarta, by taking a sample of 260 lecturers and 251 students as randomly. Descriptive analysis and t-test are used to get some conclusions. The result, lecturers exhibited a significantly higher level of Optimism and innovativeness towards using new technology than Students did. Two other dimensions, there are no significantly differences of Discomfort and Insecurity , between Lecturers and Students did.

Keywords: ICT, Academic information system, Technology Readiness.

1 Introduction

Advances in technology plays an important role in the various activities of human beings, such as business activities, trades, industries, social, cultural, not least in the field of education. Currently ICT, has been used in every activity of the organization, either partially or as a whole.

One used of ICT in higher education for the management of higher education, such as the management of new student admission, academic management, financial management, asset management, and human resources management of the college. With the presence of ICT in higher education, which is known as the Academic Information System (AIS), it is expected that the management of the college to be more effective and efficient, both in group of users which is comprised of students, faculties, parents, and other stakeholders, and in group of management of the university. To provide good service to students and faculties, XYZ University has implemented the AIS as online which managed by Bureau of Information Technology (BIT).

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 539–550, 2015.
DOI: 10.1007/978-3-662-46742-8_49

AIS specifically is designed to meet the needs of university that want a computerized educational services to improve the performance, service quality, competitiveness and quality of the human resources output. AIS is very helpful in data management of student registration process, student grades, subjects course, the data of faculties as well as the administrative faculties / departments that are still to be done manually, with the help of software will be able time effective and reduce operating costs.

AIS has also been adapted to the needs of universities including the preparation of reports of EPSBED (Evaluation Study Based on Self Evaluation Program) to the Higher Education Directorate each semester automatically. The advantages of the use of AIS, among others; user friendly, in accordance with the needs of Higher Education, suppress operating costs and Web-based / network and developed consistently.

Besides the activities of the university management, AIS are made is also useful and support for faculties and students in the learning process, research and dedication to the community. Almost all of information and a varieties of knowledge available on the Internet, and can be obtained anytime, anywhere and by anyone can get it. Lecturers and students can update their knowledge from the internet, so that science grew and evolved. AIS was built by IT technology, both in terms of hardware and software. When the implementation of AIS, however some professors or students do not use IT as well, because they are technology illiterate or they do not have funds to have the hardware and software. Or there are some of them had a little bit IT knowledge so that readiness to use IT is also little. Therefore, it is very interesting this study was conducted to see how the readiness of faculties and students to use IT technology for teaching and learning and in using AIS . The purpose of this paper is to look at one of the dimensions along which lecturers and students may differ that being their level of technology readiness. The Technology Readiness Index (TRI) was introduced by Parasuraman [1], which measures the propensity to embrace and use new technologies for accomplishing goals in home life and at work. The TRI identifies four dimensions of technology belief that impact an individual's level of techno-readiness. This paper utilizes the TRI to assess lecturer and students' willingness and likelihood to use new technology.[2] As higher education increasingly relies on technology to services delivery and enhance course offerings [3], an understanding of the differences between Lecturers and students' propensity to adopt and use innovative technology would seem to be imperative for teaching-learning service delivery of academic information system. When higher education understand these differences, they should be able to establish a collaborative and effective technology-based learning environment [2], and the strategy of AIS implementation.

2 Review Literature

2.1 Technology Readiness Index

Technology readiness (TR) refers to "people's propensity to embrace and use new technologies to accomplish goals in home life and at work" [1]. It is a combination of positive and negative technology-related beliefs. These beliefs are assumed to vary

among individuals. Collectively, these coexisting beliefs determine a person's predisposition to interact with new technology [4]. Furthermore, findings show that these beliefs can be categorized into four dimensions: optimism, innovativeness, discomfort, and insecurity [1].

- *Optimism* is defined as "a positive view of technology and a belief that offers people increased control, flexibility, and efficiency in their lives" [4]. It generally captures positive feelings about technology.
- *Innovativeness* is defined as "a tendency to be a technology pioneer and thought leader" [4] This dimension generally measures to what degree individuals perceive themselves as forefront.
- *Discomfort* is defined as "a perceived lack of control over technology and a feeling of being overwhelmed by it" [4] This dimension generally measures the fear and concerns people experience when confronted with technology.
- *Insecurity* is defined as "distrust of technology and scepticism about its ability to work properly" [4]. This dimension focuses on concerns people may have in face of technology-based transactions.

Optimism and innovativeness are drivers of technology readiness. A high score on these dimensions will increase overall technology readiness. Discomfort and insecurity, on the other hand, are inhibitors of technology readiness. Thus, a high score on these dimensions will reduce overall technology readiness [1]. Results show that the four dimensions are fairly independent, each of them making a unique contribution to an individual's technology readiness [4].

TRI emerged through an extensive multiphase research program in the United States. A replication in Great Britain has further strengthened the soundness of TRI. Tsikriktsis [5] extracted the same four-factor structure. Both studies obtained large national cross sectional samples by conducting random based telephone interviews: A total of 1000 adults (over 18 years) participated in the United States, and 400 adults (over 16 years) participated in Great Britain [1],[5].

2.2 Related Research

TRI introduced by Parasuraman has been widely used in various studies, whether TRI alone or TRI in combination with other theories or other variables in a model. TRI study in itself more likely to express the extent of a person's or organization's readiness for the technology to be used. Below, some TRI study:

1. Kuang-Ming Kuo et al [6], in their research use technology readiness to investigate nurses' personality traits in regard to technology readiness toward MEMR (mobile electronic medical record) acceptance.
2. Steven A. Taylor et al [7], in their study to assist e-Insurance marketers in developing a research-based foundation on which to make strategic decisions related to technology/Internet implementation.
3. Mu-Cheng Wu et al [8], do study in Constructing a Technology Readiness Scale for Sports Center RFID Door Security System Users

4. A.D. Berndt, S.G. Saunders & D.J. Petzer [9], in their rersearch Readiness for banking technologies in developing countries.
5. Kevin M. Elliott, Juan (Gloria) Meng [10], their study examines the influence of the Technology Readiness Index (TRI) to use self -service technology to complete retail transactions.
6. Chien-Hung Chen, Gillian Sullivan Mort [11], try to explore factors influencing technology adoption together with technology readiness, and the role of alternative technology.
7. Ahmet Emre Demirci, Nezihe Figen Ersoy [12], understanding potential customers' technology readiness and their perceptions concerning certain products and services.
8. Jonas Matthing, Per Kristensson and Anders Gustafsson [13], have aim of their paper is to explore the identification of innovative customers and the effectiveness of employing such customers to generate new service ideas in a technology-based service setting.
9. Wan Abdul Rahim Wan Mohd Isa et al [14], their study investigates the technology readiness of rural community in Malaysia, based on the gender groupings, as strategies to contribute for sustainable future of ICT-based initiatives.
10. In Manos Roumeliotis, Tsourela Maria [15], their research to Customers' technology readiness, play a lead role on predicting the perception and behavior of consumers, is tested within Greek culture and particularly higher education students.
11. Khaled A. Gad [16], in his research about Technology Readiness on Egyptians' Attitude towards E-Shopping.
12. Hossein Asgharpour b [17], has been to investigate the technology readiness of staff of a multi national chemical company operating in Iran.
13. Abdirahman Abdulahi Ahmed et al [18], in their paper proposed a model of VLE (Virtual Learning Education) readiness in higher education institutions.
14. Cathy Areu Jones [19], in his research the readiness of teacher in using the technology in teaching in the classroom.
15. Jong-Wha Lee [20], in his paper discusses the measures in building appropriate human capacities for the adaptation of new technologies in developing countries by focusing on the education strategies of East Asian economies.
16. Veronica Liljander et al [21], investigates the effects of TR on customers' (1)attitudes towards using SST(Self service technology) for airline check-in, (2) adoption of self-service check-in, and (3) evaluations of a new self service check-in on the Internet.

From all the reviews literature obtained, research on the Technology Readiness on the service delivery of the Academic Information system in higher education, is a highly relevant research and study unique.

3 Methodology

Technology readiness was assessed through the use of the 36-item Technology Readiness Index (TRI) scale developed by Parasuraman [1]. The TRI is a Likert type scale with responses ranging from Strongly Agree (5) to Strongly Disagree (1). The TRI measures an individual's propensity to adopt and use innovative technology by assessing how techno-ready individuals are. In addition, the TRI helps explains how and why different individuals adopt technology. The TRI does this by looking at both forces that attract and repel individuals away from new technology.

Data were collected using two independent convenience samples. One sample consisted of 260 Lecturers and 251 students of the XYZ university, randomly. Respondents were surveyed via a personal questionnaires.

Data analysis used descriptive analysis on a group data of lecturers and a group data of students, such as average value for each item and average value for each TRI dimension. T-test analysis will be used to look the differences between average value accros lecturer's data dan student's data.

4 Result and Discussion

4.1 TRI Dimension between Lecturers Vs Students

Table 1 shows the results of t-tests for assessing differences in mean scores between lecturers and students across the four dimensions comprising the TRI. Lecturers exhibited a significantly higher level of Optimism towards using new technology than did Students (3.92 vs. 3.62). Therefore, lecturers appear to have a stronger belief than students that technology can benefit their lives, as well as give them more control over their life. The results in Table 3 also show that lecturers reported a significantly higher level of Innovativeness related to the propensity to use new technology than students did (3.85 vs. 3.63). This finding demonstrates that lecturers exhibit a greater desire to experiment with new technologies than students do.

However, reported there are no significantly differences of Discomfort towards using new technology between lecturers and students did. So that, both exhibited a small feeling of lack of control over technology and the confidence in making the technology work. Also reported there are no significantly differences of Insecurity towards using new technology between lecturers and students. Both appear to feel a small need for assurance that a technology-based product will operate reliably and accurately .

4.2 Individual TR Items – (Lecturers vs. Students)

Given that significant differences were found between lecturers and students across all four TRI dimensions, additional t-tests were conducted on the 36 individual items comprising the TRI scale in an effort to further assess differences between lecturers and students with respect to their propensity to embrace and use new technology. The

Table 1. Average value and t-test result of each Dimension

Optimism					
	CODE	Mean	Std. Deviation	t	Sig. (2-tailed)
OPT	Lecturer	3,91500	,183379	2,791	,012
	Student	3,62400	,274032	2,791	,013
Innovativeness					
	CODE	Mean	Std. Deviation	t	Sig. (2-tailed)
INNOV	Lecturer	3,85429	,098634	2,604	,023
	Student	3,62857	,206997	2,604	,030
Discomfort					
	CODE	Mean	Std. Deviation	t	Sig. (2-tailed)
DISCOM	Lecturer	3,43300	,086545	1,855	,080
	Student	3,33700	,138888	1,855	,083
Insecurity					
	CODE	Mean	Std. Deviation	t	Sig. (2-tailed)
INSEC	Lecturer	3,29556	,122893	1,506	,152
	Student	3,21111	,114831	1,506	,152

results are presented in Table 2. lecturer reported significantly higher mean scores on 8 of the 10 individual items that comprise the Optimism dimension. Lecturers felt stronger than students at:OP1, OP2, OP3, OP4, OP5, OP7, OP9, and OP10. But lecturers and student no differences at item OP6 and OP8 .

Table 2. Average value and t-test result of Optimism Dimension

		CODE	Mean	Std. Deviation	t	Sig. (2-tailed)
OP1	Technology gives people more control over their daily lives	Lecturer	3,91	,733	12,082	,000
		Student	3,08			
OP2	Products/services that use new technologies are convenient to use	Lecturer	3,63	,842	2,823	,005
		Student	3,43			
OP3	I like the idea of doing business via computers because I'm not limited to regular business hours.	Lecturer	3,63	,919	2,581	,010
		Student	3,42			
OP4	I prefer to use the most advanced technology available.	Lecturer	4,12	,872	5,816	,000
		Student	3,73			
OP5	I like computer programs that allow me to tailor things to fit my own needs.	Lecturer	4,12	,945	4,787	,000
		Student	3,75			

OP6	Technology makes me more efficient in my occupation	Lecturer	3,90	,941	,043	,966
		Student	3,90			
OP7	I find new technologies to be mentally stimulating.	Lecturer	4,04	,967	4,502	,000
		Student	3,68			
OP8	Technology gives me more freedom of mobility.	Lecturer	4,07	,777	,024	,981
		Student	4,06			
OP9	Learning technology can be as rewarding as the technology itself.	Lecturer	3,94	,943	3,955	,000
		Student	3,63			
OP10	I feel confident that machines will follow through with what I instructed them to do.	Lecturer	3,79	,880	3,074	,002
		Student	3,56			

Lecturers also reported at table-3 significantly higher mean scores on 4 out of 7 individual items that reflect the Innovative dimension, which assesses an individual's desire to experiment with new technologies. Lecturers reported stronger agreement to statements such in IN1, IN2, IN4 and IN6. The rest item in Innovativeness dimension, no differences significanly between lecturers and students.

Table 3. Average value and t-test result of Innovativeness Dimension

		CODE	Mean	Std. Deviation	t	Sig. (2-tailed)
IN1	Other people come to me for advice on new technologies	Lecturer	3,97	,792	7,282	,000
		Student	3,50			
IN2	It seems my friends are learning more about the newest technologies than I am.	Lecturer	3,93	,797	6,788	,000
		Student	3,49			
IN3	In general, I am among the first in my circle of friends to acquire new technology when it appears.	Lecturer	3,95	,641	-,181	,856
		Student	3,96			
IN4	I can usually figure out new high-tech products and services without help from others.	Lecturer	3,84	,730	6,805	,000
		Student	3,43			
IN5	I keep up with the latest technological developments in my areas of interest.	Lecturer	3,81	,919	-,114	,910
		Student	3,82			
IN6	I enjoy the challenge of figuring out high-tech gadgets.	Lecturer	3,77	,811	3,784	,000
		Student	3,47			
IN7	I find I have fewer problems than other people in making technology work for me.	Lecturer	3,71	,890	-,222	,824
		Student	3,73			

In table-4, Lecturers, however, reported significantly higher mean scores than students on 4 out of 10 individual items that reflect the Discomfort dimension, those ara: DISC2, DISC4, DISC6, and DISC 7, which assesses an individual's feeling of

lacking both control over technology and the confidence in making the technology work. The rest item in Disfomfort dimension, no differences significanly between lecturers and students.

Table 4. Average value and t-test result of Discomfort Dimension

		CODE	Mean	Std. Deviation	t	Sig. (2-tailed)
DIS1	Technical support lines are not helpful because they don't explain things in terms I understand.	Lecturer	3,33	,889	1,942	,053
		Student	3,17			
DIS2	Sometimes, I think that technology systems are not designed for use by ordinary people.	Lecturer	3,45	,894	2,292	,022
		Student	3,25			
DIS3	There is no such thing as a manual for a high-tech product or service that's written in plain language.	Lecturer	3,45	1,065	,084	,933
		Student	3,44			
DIS4	When I get technical support from a provider of a high-tech product or service, I sometimes feel as if I am taken advantage of by someone who knows more than I do.	Lecturer	3,54	,922	2,038	,042
		Student	3,37			
DIS5	If I buy a high-tech product or service, I prefer to have the basic model over one with a lot of extra features.	Lecturer	3,39	,937	,073	,942
		Student	3,38			
DIS6	It is embarrassing when I have trouble with a high-tech gadget while people are watching.	Lecturer	3,38	,953	1,977	,049
		Student	3,21			
DIS7	There should be caution in replacing important people-tasks with technology because new technology can break down or get disconnected	Lecturer	3,35	,932	2,331	,020
		Student	3,16			
DIS8	Many new technologies have health or safety risks that are not discovered until after people have used them.	Lecturer	3,39	1,036	,196	,845
		Student	3,37			
DIS9	New technology makes it too easy for governments and companies to spy on people.	Lecturer	3,61	,926	,170	,865
		Student	3,60			
DIS10	Technology always seems to fail at the worst possible time.	Lecturer	3,44	1,057	,216	,829
		Student	3,42			

In table-5, Lecturers also reported significantly higher mean scores than students did on 2 out of 10 individual items that comprise the Insecurity dimension at INS6 and INS9. This dimension measures an individual's need for assurance that a technology-based product will operate reliably and accurately. The rest item in Insecurity dimension, no differences significanly between lecturers and students.

Table 5. Average value and t-test result of Insecurity Dimension

		CODE	Mean	Std. Deviation	t	Sig. (2-tailed)
INS1	The human touch is very important when doing business with a company.	Lecturer	3,27	,908	-,300	,764
		Student	3,29			
INS2	When I call a business, I prefer to talk to a person rather than a machine.	Lecturer	3,18	1,022	,281	,779
		Student	3,16			
INS3	If I provide information to a machine or over the Internet, I can never be sure it really gets to the right place.	Lecturer	3,07	,877	-,547	,585
		Student	3,12			
INS4	I do not consider it safe giving out a credit card number over a computer.	Lecturer	3,29	1,028	,237	,813
		Student	3,27			
INS5	I do not consider it safe to do any kind of financial business online.	Lecturer	3,42	,886	1,198	,232
		Student	3,31			
INS6	I worry that information I send over the Internet will be seen by other people	Lecturer	3,31	,815	2,615	,009
		Student	3,09			
INS7	I do not feel confident doing business with a place that can only be reached online.	Lecturer	3,37	,926	1,326	,185
		Student	3,26			
INS8	Any business transaction I do electronically should be confirmed later with something in writing.	Lecturer	3,48	,952	1,189	,235
		Student	3,37			
INS9	Whenever something gets automated, I need to check carefully that the machine or computer is not making mistakes	Lecturer	3,27	,834	2,887	,004
		Student	3,03			

The findings of this study show that students exhibit a lower propensity to embrace and use new technology than lecturers do. As the implications of the results of this study at XYZ university, for teaching and learning activities, faculties and students should be encouraged to use AIS and Internet such as using lecture material from the Internet and interact through the internet like giving assignments to students, giving lectures and discussions. Similarly, students with e-mail can send tasks to the lecturers. With the Discomfort dimensions are quite high (> 3), then the university must convince faculties and students that the AIS system is controlled everything and

AIS is working well for management purposes as well as teaching and learning purposes. For example with AIS, attendance faculties, students and employees are more controlled, course selection and faculties will be able to set up the classrooms, student registration can be done online, recording faculties workload each semester becomes more controlled etc. Likewise, the presence of Insecurity dimension scores were greater than 3, the university must be assured that the AIS system works well and accurately, so that all transactions conducted by faculties and students are safe and can be guaranteed, such as the entered of grades can not be changed by anyone, money posted by students is up to the financial departement, optional subjects students can not be changed by anyone etc. To assure that, the university can do through refreshing about the use of AIS's once every few months, or put up banners on the benefits and how to use of AIS. To new students always done socialization of AIS as well.

5 Conclusions

The purpose of this study was to assess differences in lecturers and students with respect to their level of technology readiness. The findings of this study show that students exhibit a lower propensity to embrace and use new technology than do lecturers. Lecturers exhibited a significantly higher level of Optimism and innovativeness towards using new technology than Students did. Two other dimensions, there are no significantly differences of Discomfort and Insecurity , between lecturers and students did. In each item, lecturer better in 18 items than student while the rest18 item no indicate significant differences. An understanding of the differences between lecturer and students' propensity to adopt and use innovative technology is an important step in creating teaching-learning competency in technology-based learning environments and in implementation of Academic Information System as good and right. Institution must understand these differences in order to establish a effective teaching-learning strategy and strategy of service delivery by using AIS.

References

1. Parasuraman, A.: Technology readiness index (TRI): A multiple-item scale to measure readiness to embrace new technologies. Journal of Service Research 2, 307–320 (2000), doi:10.1177/109467050024001
2. Elliott, K.M., Hall, M.C., Meng, J.(G.): Student Technology Readiness and its Impact on Cultural Competency. College Teaching Methods & Styles Journal 4(6) (June 2008)
3. ECAR.:The ECAR study of undergraduate students and information technology (Study 6), EDUCAUSE Center for Applied Research (2007)
4. Parasuraman, A., Colby, C.L.: Techno-ready marketing: How and why your customers adopt technology. Free Press, New York (2001)
5. Tsikriktsis, N.: A technology readiness-based taxonomy of costumers: A replication and extension. Journal of Service Research 7, 42–52 (2004), doi:10.1177/1094670504266132
6. Kuo, K.-M., Liu, C.-F., Ma, C.-C.: An investigation of the effect of nurses'technology readiness on the acceptance of mobile electronic medical record systems (2013),

http://www.biomedcentral.com/1472-6947/13/88 (accessed December 4, 2014)
7. Taylor, S.A., Celuch, K., Goodwin, S.: Technology Readinessin the e-Insurance Industry:An Exploratory Investigation and Development of an Agent Technology e-Consumption Model (2014), http://www.insuranceissues.org/PDFs/252TCG.pdf (accessed December 4, 2014)
8. Wu, M.-C., Chen, C.-C., Tseng, C.-H.: Constructing a Technology Readiness Scale for Sports Center RFID Door Security System Users (2014), http://www.jgbm.org/page/2%20Chih-Hung%20Tseng.pdf (accessed December 4, 2014)
9. Berndt, A.D., Saunders, S.G., Petzer, D.J.: Readiness for banking technologies in developing countries (2014), http://www.ajol.info/index.php/sabr/article/viewFile/76386/66843 (accessed December 4, 2014)
10. Elliott, K.M., Meng, J.(G.).: Assessing Chinese Consumers' Likelihood To Adopt Self-Service Technologies. International Business & Economics Research Journal 8(2) (2009), http://www.cluteinstitute.com/ojs/index.php/IBER (accessed December 4, 2014)
11. Chen, C.-H., Mort, G.S.: Ready or not? That is the Question for Consumer Technology Acceptance (2014), http://www.researchgate.net/publication/260060943_Ready_or_not_That_is_thequestion_for_Consumer_Technology_Acceptance (accessed December 4, 2014)
12. Demirci, A.E., Ersoy, N.F.: Technology Readiness for Innovative High-Tech Products: How Consumers Perceive and Adopt New Technologies (2014), http://isletme.anadolu.edu.tr/media/articles/demirci_ersoy01.pdf (accessed December 4, 2014)
13. Matthing, J., Kristensson, P., Gustafsson, A.: Developing successful technology-based services: the issue of identifying and involving innovative users (2014), http://www.emeraldinsight.com/doi/abs/10.1108/08876040610679909 (accessed December 4, 2014)
14. Rahim, W.A., Isa, W.M., et al.: Rural Informatics Engineering: Assessing Technology Readiness of Rural Community towards Implementation of e-learning (2014), http://www.wseas.us/e-library/conferences/2013/Malaysia/EDUETE/EDUETE-02.pdf (accessed December 4, 2014)
15. Roumeliotis, M., Maria, T.: Perception And Adoption of Technology Based Services by Students of Higher education. International Journal of Scientific and Research Publications 4(3) (March 2014) ISSN 2250-3153, http://www.ijsrp.org/research-paper-0314/ijsrp-p27126.pdf (accessed December 4, 2014)
16. Gad, K.A.: The Impact of Demographics and Technology Readiness on Egyptians' Attitude towards E-Shopping. In: International Conference on Transport, Environment & Medical Sciences (ICTEMS 2012), Kuala Lumpur, Malaysia, August 25-26 (2012), http://psrcentre.org/images/extraimages/812592.pdf (accessed December 4, 2014)
17. Asgharpour, H.: Measuring The Staff Technology Readiness, The case a Multi National Chemical Company Operating in Iran. Master's Thesis. Tarbiat Modares University Faculty of Engineering, Department Industrial Engineering Lulea Uni versity of Tchnology (2006), http://epubl.ltu.se/1653-0187/2007/050/index-en.html (accessed December 4, 2014)

18. Ahmed, A.A., et al.: An Evaluation of Virtual Learning Environment Readiness in Higher Education Institutions (HEIs). Journal of Information Systewms Research and Innovation (2014), http://seminar.utmspace.edu.my/jisri/download/F_FinalPublished/Pub11VirtualLearningReadinessinHEIs_amend.pdf (accessed December 4, 2014)
19. Jones, C.A.: Preparing Teachers to Use Technology Educators are urged to incorporate technology into instruction, but the effectiveness of educational technology is determined by teachers' readiness to use it, not by its mere presence in the classroom (2014), https://www.nassp.org/portals/0/content/48158.pdf (accessed December 4, 2014)
20. Lee, J.-W.: Education for Technology Readiness: Prospects for Developing Countries. Journal of Human Development 2(1) (2001), http://unpan1.un.org/intradoc/groups/public/documents/apcity/unpan002133.pdf (accessed December 4, 2014)
21. Liljander, V., et al.: Technology readiness and the evaluation and adoption of self-service technologies. Journal of Retailing and Consumer Services 13, 177–191 (2006), https://www.deepdyve.com/lp/elsevier/technology-readiness-and-the-evaluation-and-adoption-of-self-service-8GzNknL5F0 (accessed December 4, 2014)

Tool Support for Cascading Style Sheets' Complexity Metrics

Adewole Adewumi[1], Onyeka Emebo[1], Sanjay Misra[2], and Luis Fernandez[3]

[1] Department of Computer and Information Sciences, Covenant University, OTA, Nigeria
[2] Department of Computer Engineering, Covenant University, OTA Nigeria
[3] Department of Computer Science, University of Alcala, Madrid, Spain
{wole.adewumi,onye.emebo}@covenantuniversity.edu.ng,
smisra@atilim.edu.tr,luis.fernandezs@uah.es

Abstract. Tools are the fundamental requirement for acceptability of any metrics programme in the software industry. It is observed that majority of the metrics proposed and are available in the literature lack tool support. This is one of the reasons why they are not widely accepted by the practitioners. In order to improve the acceptability of proposed metrics among software engineers that develop Web applications, there is need to automate the process. In this paper, we have developed a tool for computing metrics for Cascading Style Sheets (CSS) and named it as CSS Analyzer (CSSA). The tool is capable of measuring different metrics, which are the representation of different quality attributes: which include understandability, reliability and maintainability based on some previously proposed metrics. The tool was evaluated by comparing its result on 40 cascading style sheets with results gotten by the manual process of computing the complexities. The results show that the tool computes in far less time when compared to the manual process and is 51.25% accurate.

Keywords: Cognitive complexity, CSS Metrics, Tool Support.

1 Introduction

In recent times, Cascading Style Sheets (CSS) have become indispensable in the development of Web applications. They can be used for styling eXtensible Markup Language (XML) or HyperText Markup Language (HTML) documents. To style an HTML document, CSS can be applied in three ways namely: by placing them within the <head> tags of HTML; by placing them as tag attributes within the various other HTML tags that can be contained within the <body> tag of an HTML document; or by creating them as a separate document with extension (.css) and linking them to the <head> section of the HTML document. The latter of these procedures is a better practice as it separates content from presentation thereby promoting maintainability. Since CSS are an integral part of Web applications, they also add to the increasing complexity of such applications. In a previous work [1], we presented a suite of metrics to measure complexity in CSS at the International Conference on Computational Science and Its Applications (ICCSA 2012). Prior to this, not much had been done in

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 551–560, 2015.
DOI: 10.1007/978-3-662-46742-8_50

this regard. The metrics were based on cognitive complexity and were computed manually. This is a slow and tedious process that makes it of no practical use to software engineers and Web developers. Though we carried out a preliminary evaluation of the metrics and found them to be well structured, we proposed as future work to provide a tool to simplify the measurement process. This informs the motivation for this current paper.

Tool writing according to [2] is becoming a forgotten art. This is evidenced today by a growing body of literature that consists mainly of proposed and validated complexity metrics with no tools to measure them. Table I shows an in exhaustive list of such literature. In the XML schema document (XSD) domain, a metric has been proposed to measure complexity in XML schema documents (XSDs) [9]. The metric does this by considering the internal building blocks of XSD. The metric was demonstrated with examples and performed well in comparison to similar measures. To compute the metric however, no tool was provided. Similar to this, a design complexity metric was proposed for XSD in [17]. This metric covers all the major factors that affect the complexity of XSD. In addition, due to the diversity in structure of W3C XML schema languages, a metric based on the concept of entropy from information theory was proposed in [23] for assessing the structural complexity of XSDs. As for XSDs written in W3C Document Type Definition (DTD) language, a metric also based on the entropy concept from information theory was proposed in [14] to measure the structural complexity of XSDs written in DTD. The work was extended in [24] to include Distinct Structured Element Repetition Scale (DSERS) metric, which also measures the structural complexity of schemas in DTD language. This metric exploits a directed graph representation of a schema document and considers the complexity of schema due to its similar structured elements and the occurrences of these elements. In the XML/Web services domain, a data complexity metric for XML Web services was proposed in [16] which assesses the quality of Web services in terms of maintainability. Similarly, a suite of metrics for XML Web services was also proposed in [21] which includes: data weight of a web service description language, distinct message ratio metric, message entropy metric and message repetition scale metric. All the proposed metrics in the suite were evaluated theoretically and validated empirically. A comparative study with similar measures also proved the worth of the metric suite. In the coding language domain, a complexity metric was proposed in [20] for evaluating object-oriented code with emphasis on Python, Java and C++. The metric was validated empirically on real projects but no tool was developed for the metric. Similarly, a complexity metric was proposed in [19] called JavaScript Cognitive Complexity Measure (JCCM) for measuring the complexity of JavaScript code. Again the metric was evaluated theoretically and validated empirically but no tool support was made available. Other metrics based mostly based on cognitive informatics include: Modified Cognitive Complexity Measure [4], [5]; Complexity Measure based on Cognitive Weights [6]; Cognitive Program Complexity Measure [7]; Object Oriented Complexity Metric Based on Cognitive Weights [8]; a New Complexity Metric Based on Cognitive Informatics [10]; Object Oriented Programs Complexity Measure [11]; Unique Complexity Metric [12]; Weighted Class Complexity [13]; Unified Complexity Measure [18]; and Inheritance Complexity Metric for Object-Oriented Code [22]. In all, these measures have no tool support.

Since tools are a fundamental requirement for the acceptance of any metric in the software industry, we are extending our previous work and developing a tool for computing metrics of CSS. The tool is named CSS Analyzer (CSSA). The rest of the paper is organized as follows: Section 2 gives a description of the tool developed. Section 3 evaluates the tool by comparing its result with the manual approach to measurement. Section 4 discusses the results of the comparison while Section 5 concludes the paper.

2 Description of CSSA

CSSA was developed using the Java programming language. In this section, we describe CSSA based on the metrics that it measures.

2.1 Rule Length (RL)

CSS is made up of rules. This metric counts the number of lines of rules in a CSS without taking into account white spaces or comment lines [1]. The pseudo code used to implement this functionality in CSSA is given as:

```
Read a CSS file;
Initialize a line counter variable to zero;
While not End of CSS File
  If a line is not empty and is not a comment;
    Increment line counter variable by 1;
```

Table 1. Metrics proposed having no tool support

S/N	Metrics name	Reference
1	Cognitive Complexity Measure	[3]
2	Modified Cognitive Complexity Measure	[4], [5]
3	Complexity Measure Based on Cognitive Weights	[6]
4	Cognitive Program Complexity Measure	[7]
5	Object Oriented Complexity Metric Based on Cognitive Weights	[8]
6	Complexity Metric for XML Schema Documents	[9]
7	A New Complexity Metric Based on Cognitive Informatics	[10]
8	Object Oriented Programs Complexity Measure	[11]
9	Unique Complexity Metric	[12]
10	Weighted Class Complexity	[13]
11	Entropy Metric for XML DTD Documents	[14]
12	Cognitive Functional Sizes	[15]
13	Data Complexity Metrics for XML Web Services	[16]
14	Design Complexity Metric for XML Schema Documents	[17]
15	Unified Complexity Measure	[18]
16	Complexity Metric for JavaScript	[19]
17	Python Language Complexity Metric	[20]
18	Metrics suite for maintainability of XML Web Services	[21]
19	Inheritance Complexity Metric for Object-Oriented Code	[22]
20	Entropy of XML Schema Document	[23]
21	DTD Metrics	[24]

2.2 Number of Rule Blocks

A rule block in CSS refers to a selector and its attributes depicted as:

```
/* Syntax of a rule block */
Selector [, selector2, ...] [:pseudo-class] {
  Property: value;
  [Property2: value2;
...]
}
```

The pseudo code used to implement this functionality in CSSA is given as:

```
Read a CSS file;
Initialize a brace counter variable to zero;
While not End of CSS File
  Increment brace counter variable by 1 every time an
open brace is read
```

2.3 Number of Attributes Defined per Rule Block (NADRB)

NADRB as defined in [1] determines the average number of attributes defined in the rule blocks of a CSS file. The formula for calculating it is given in (1):

$$\text{NADRB} = \Sigma \text{ rule\_block\_attributes} / \Sigma \text{ rule\_blocks} \quad (\mathbf{1})$$

The pseudo code for computing this metric in CSSA is given as:

```
Read a CSS file;
Initialize a semi colon counter variable to zero;
Initialize a close brace counter variable to zero;
While not End of CSS File
  Increment close brace counter variable by 1 every time
an open brace is encountered
  Increment semi colon counter variable by 1 every time a
semi colon is encountered
Divide the semi colon counter by the close brace counter
to get the NADRB value
```

2.4 Number of Cohesive Rule Blocks (NCRB)

NCRB measures the number of rule blocks in a CSS file possessing a single attribute. The pseudo code for computing this metric in CSSA is given as:

```
Read a CSS file;
Initialize counter variable to zero;
While not End of CSS File
```

```
If the number of semi colons within a rule block is one
  Increment counter variable by 1;
Else
  Do nothing;
```

3 Evaluation of the Tool

In this section, we present the evaluation of the tool by applying it on forty real CSS files downloaded from the Internet. We compare the results obtained for each metric with the results gotten by manual computation of the metrics. Table II presents the CSS Ids and the web links from which they were downloaded. Table III shows a comparison of the results obtained by applying CSSA to the CSS files as well as the results obtained by manual computation.

Table 2. CSS IDs and Web Links

CSS ID	Web Link
1	http://www.freecsstemplates.org/download/zip/argon
2	http://www.freecsstemplates.org/download/zip/boldness
3	http://www.freecsstemplates.org/download/zip/bolness2
4	http://www.freecsstemplates.org/download/zip/classifieds
5	http://www.freecsstemplates.org/download/zip/combinations
6	http://www.freecsstemplates.org/download/zip/compass
7	http://www.freecsstemplates.org/download/zip/consistent
8	http://www.freecsstemplates.org/download/zip/corporatestuff
9	http://www.freecsstemplates.org/download/zip/estatebroker
10	http://www.freecsstemplates.org/download/zip/flamingo
11	http://www.freecsstemplates.org/download/zip/flowering
12	http://www.freecsstemplates.org/download/zip/fotofolium
13	http://www.freecsstemplates.org/download/zip/fruityblue
14	http://www.freecsstemplates.org/download/zip/handcrafted
15	http://www.freecsstemplates.org/download/zip/igunalounge
16	http://www.freecsstemplates.org/download/zip/infrastructure
17	http://www.freecsstemplates.org/download/zip/inwild
18	http://www.freecsstemplates.org/download/zip/islandpalm
19	http://www.freecsstemplates.org/download/zip/lettering
20	http://www.freecsstemplates.org/download/zip/limitless
21	http://www.freecsstemplates.org/download/zip/networked
22	http://www.freecsstemplates.org/download/zip/officememo
23	http://www.freecsstemplates.org/download/zip/outdoor
24	http://www.freecsstemplates.org/download/zip/petalsandflowers
25	http://www.freecsstemplates.org/download/zip/redallover
26	http://www.freecsstemplates.org/download/zip/redandblack
27	http://www.freecsstemplates.org/download/zip/reinstated
28	http://www.freecsstemplates.org/download/zip/rifle

Table 2. (*Continued*)

29	http://www.freecsstemplates.org/download/zip/simplified
30	http://www.freecsstemplates.org/download/zip/woodcrafting
31	http://www.freecsstemplates.org/download/zip/stampalike
32	http://www.freecsstemplates.org/download/zip/naturalprime
33	http://www.freecsstemplates.org/download/zip/grasstown
34	http://www.freecsstemplates.org/download/zip/fullycharge
35	http://www.freecsstemplates.org/download/zip/spikyflower
36	http://www.freecsstemplates.org/download/zip/simpledisplay
37	http://www.freecsstemplates.org/download/zip/modelling
38	http://www.freecsstemplates.org/download/zip/surround
39	http://www.freecsstemplates.org/download/zip/halcyonic
40	http://www.freecsstemplates.org/download/zip/bigbusiness2

Figs. 1 - 4 give a graphical comparison of the results gotten from the tool as well as that computed by hand. The next section discusses this in detail.

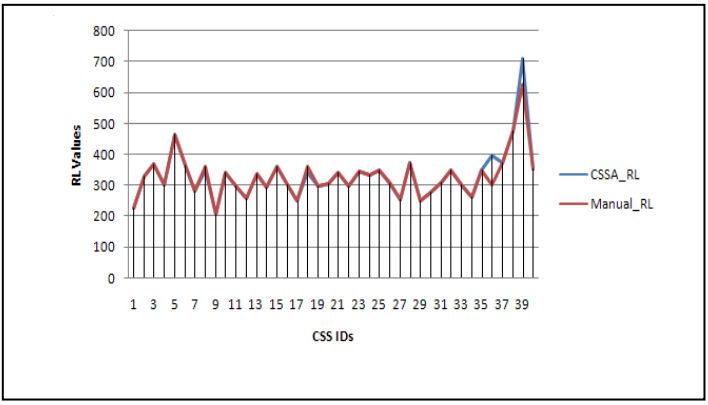


Fig. 1. This figure shows the correlation between Rule Length values calculated by hand and using the tool

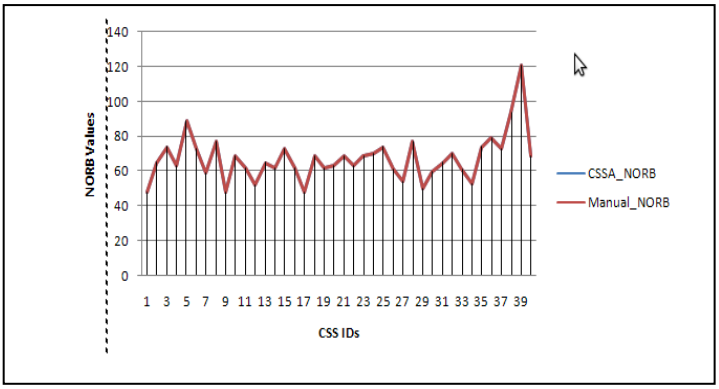


Fig. 2. This figure shows the correlation between NORB values calculated by hand and using the tool

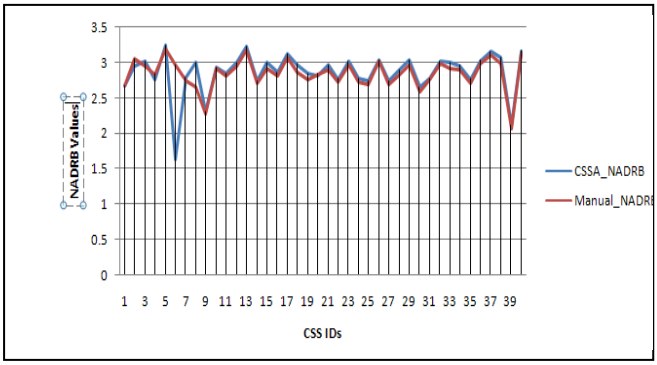


Fig. 3. This figure shows the correlation between NADRB values calculated by hand and using the tool

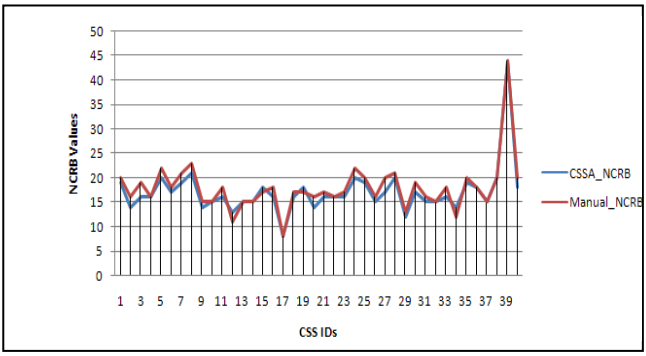


Fig. 4. This figure shows the correlation between NCRB values calculated by hand and using the tool

Table 3. Comparison between results obtained using CSSA and manual computation

ID	CSSA				MANUALLY			
	RL	NORB	NADRB	NCRB	RL	NORB	NADRB	NCRB
1	224	48	2.68	19	224	48	2.67	20
2	329	65	2.95	14	329	65	3.06	16
3	368	74	3.02	16	368	74	2.96	19
4	300	63	2.76	16	300	63	2.84	16
5	463	89	3.25	20	463	89	3.20	22
6	363	73	1.63	17	364	73	2.97	18
7	280	59	2.79	19	280	59	2.75	21
8	349	77	3.00	21	362	77	2.66	23
9	205	48	2.31	14	205	48	2.27	15
10	340	69	2.94	15	340	69	2.93	15
11	298	62	2.85	16	298	62	2.81	18
12	258	52	3.01	13	258	52	2.96	11

Table 3. (*Continued*)

13	336	65	3.23	15	336	65	3.19	15
14	292	62	2.75	15	292	62	2.71	15
15	359	73	3.00	18	359	73	2.92	17
16	299	62	2.87	16	300	62	2.82	18
17	248	48	3.12	8	248	48	3.08	8
18	340	69	2.98	16	360	69	2.87	17
19	296	62	2.85	18	296	62	2.77	17
20	305	63	2.82	14	305	63	2.84	16
21	339	69	2.97	16	339	69	2.91	17
22	298	63	2.77	16	298	63	2.73	16
23	343	69	3.02	16	343	69	2.97	17
24	332	70	2.78	20	332	70	2.73	22
25	348	74	2.75	19	348	74	2.70	20
26	306	61	3.04	15	307	61	3.02	16
27	253	54	2.74	17	253	54	2.69	20
28	373	77	2.90	20	373	77	2.84	21
29	249	50	3.04	12	249	50	2.98	13
30	276	60	2.65	17	276	60	2.60	19
31	309	65	2.78	15	309	65	2.77	16
32	349	70	3.02	15	349	70	2.99	15
33	302	61	3.00	16	302	61	2.92	18
34	260	53	2.96	14	260	53	2.91	12
35	349	74	2.77	19	349	74	2.72	20
36	395	79	3.03	18	302	79	3.00	18
37	374	73	3.16	15	374	73	3.11	15
38	475	94	3.07	20	475	94	3.00	20
39	711	121	2.11	44	628	121	2.07	44
40	354	69	3.17	18	354	69	3.13	20

4 Discussion

The comparison in Table III shows that for the RL metric only 7 (17.5%) of the CSS files analyzed do not give same results for the manual and automated complexity computation process. In other words, 82.5% of the CSS files match when computed using both automated and manual process. It is easy to identify the files that do not match by looking at the chart in Fig. 1. For instance, CSS file with ID 36 has an RL value of 395 when the tool is used but when calculated by hand the value obtained is 302 a sharp difference.

For the NORB metric, we observe that the results obtained by the tool and by manual computation are exactly alike for all 40 (100%) CSS files analyzed. This explains why the graph in Fig. 2 seems to have only one (red) color. A direct opposite of this is the case of NADRB metric where none of the 40 (0%) CSS files analyzed give same results for the manual and automated computation process. This is supported by the irregular lines seen in Fig. 3 notable among them is the sharp difference between the automated and manual computation for CSS ID 6 (1.63 and 2.97 respectively).

As for the NCRB metric, only 9 (22.5%) out of the 40 CSS files analyzed have the same results. The differences in value for both automated and manual computation is not significant as can be seen in Fig. 4. To this end, summing the percentage match for each metric and dividing the total by the number of metrics considered helps to determine the percentage accuracy of the tool given as:

$(82.5 + 100 + 0 + 22.5)/4 = 51.25\%$ accuracy

The tool takes an average of 11 seconds to compute all four metrics while a manual computation takes an average of 390 seconds. This implies that the tool is 35 times faster than the manual computation process.

5 Conclusion

This paper presented the description and evaluation of CSSA - a tool for computing RL, NORB, NADRB and NCRB metrics for a CSS file. CSSA was implemented to alleviate the cumbersome process of determining the complexity of CSS by hand. The result from the comparison to the manual computation approach shows that CSSA is 35 times faster with an accuracy of 51.25%. Web developers and engineers can utilize the tool in its current form. As future work, we intend to improve on the accuracy of CSSA and also extend it to compute entropy of CSS files.

References

1. Adewumi, A., Misra, S., Ikhu-Omoregbe, N.: Complexity Metrics for Cascading Style Sheets. In: Murgante, B., Gervasi, O., Misra, S., Nedjah, N., Rocha, A.M.A.C., Taniar, D., Apduhan, B.O. (eds.) ICCSA 2012, Part IV. LNCS, vol. 7336, pp. 248–257. Springer, Heidelberg (2012)
2. Spinellis, D.: Tool writing: a forgotten art? IEEE Software 22, 9–11 (2005)
3. Misra, S., Misra, A.K.: Evaluating cognitive complexity measure with Weyuker Properties. In: Proc. 3rd IEEE International Conference on Cognitive Informatics, Victoria, Canada, pp. 103–108 (2004)
4. Misra, S.: Modified cognitive complexity measure. In: 21st International Symposium on Computer and Information Sciences, pp. 1050–1059 (2006)
5. Misra, S.: Validating modified cognitive complexity measure. ACM SIGSOFT Software Engineering Notes 32, 1–5 (2007)
6. Misra, S.: Complexity measure based on cognitive weights. International Journal of Theoretical and Applied Computer Sciences 1, 1–10 (2006)
7. Misra, S.: Cognitive program complexity measure. In: Proc. 6th IEEE Int. Conf. on Cognitive Informatics, Lake Tahoe, CA, pp. 120–125 (2007)
8. Misra, S.: An object oriented complexity metric based on cognitive weights. In: Proc. 6th IEEE Int. Conf. on Cognitive Informatics, Lake Tahoe, CA, pp. 134–139 (2007)
9. Basci, D., Misra, S.: Complexity metric for XML schema documents. In: Proc. International Conference on SOA and Web services, pp. 1–14 (2012)
10. Misra, S., Akman, I.: A New Complexity Metric Based on Cognitive Informatics. In: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (eds.) RSKT 2008. LNCS (LNAI), vol. 5009, pp. 620–627. Springer, Heidelberg (2008)

11. Misra, S., Akman, I.: Measuring complexity of object oriented programs. In: Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2008, Part II. LNCS, vol. 5073, pp. 652–667. Springer, Heidelberg (2008)
12. Misra, S., Akman, I.: A unique complexity metric. In: Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., Gavrilova, M.L. (eds.) ICCSA 2008, Part II. LNCS, vol. 5073, pp. 641–651. Springer, Heidelberg (2008)
13. Misra, S., Akman, I.K.: Weighted class complexity: a measure of complexity for object oriented system. Journal of Information Science and Engineering 24, 1689–1708 (2008)
14. Basci, D., Misra, S.: Entropy metric for XML DTD documents. ACM SIGSOFT Software Engineering Notes 33, 1–6 (2008)
15. Misra, S.: Cognitive functional sizes. Int. J. of Software Science and Computational Intelligence, IGI-Global (2009)
16. Basci, D., Misra, S.: Data complexity metrics for XML Web Services. Advances in Electrical and Computer Engineering 9, 9–15 (2009)
17. Basci, D., Misra, S.: Measuring and evaluating a design complexity metric for XML schema documents. Journal of Info. Sci. and Engineering 25, 1405–1425 (2009)
18. Misra, S., Akman, I.: Unified complexity measure: a measure of complexity. Proc. Nat. Acad. Sci., 167–176 (2010)
19. Misra, S., Cafer, F.: Estimating quality of JavaScript. The International Arab Journal of Information Technology 9, 535–543 (2012)
20. Misra, S., Cafer, F.: Estimating complexity of programs in Python Language. Technical Gazette 18, 23–32 (2011)
21. Basci, D., Misra, S.: Metrics suite for maintainability of eXtensible markup language Web services. IET Software 5, 1–22 (2011)
22. Misra, S., Akman, I., Koyuncu, M.: An inheritance complexity metric for object-oriented code: A cognitive approach. Sadhana 36, 317–337 (2011)
23. Basci, D., Misra, S.: Entropy as a measure of quality of XML schema document. The International Arab Journal of Information Technology 8, 75–83 (2011)
24. Basci, D., Misra, S.: Document Type Definition (DTD) metrics. Romanian Journal of Info. Science and Technology 14, 31–50 (2011)

Intelligent Systems for Enterprise, Government and Society

Generic Quantitative Assessment Model for Enterprise Resource Planning (ERP) System

Olivia and Kridanto Surendro

School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
olivia.tan@students.itb.ac.id,
surendro@gmail.com

Abstract. Enterprise resource planning (ERP) system has been proved to be a powerful for managing and integrating enterprise. Thus, the number of ERP implementation is increasing every year. However, some enterprises, especially the ones implementing such large system for the first time, are facing difficulties in choosing the suitable type of ERP system. A lot of ERP system types are available in the market, making enterprises hard to choose the one that is better and more suitable for them. Available approach in ERP selection mostly only cover qualitative analysis. The ones with quantitative support require proper knowledge in technology and statistics, making it hard to be directly used in enterprises. This paper addresses a general framework for assessing ERP system quantitatively, based on its selection criteria and critical success factors (CSFs) in implementation.

Keywords: Enterprise resource planning, Software development management, project engineering.

1 Introduction

Effective use of enterprise resource planning (ERP) system is one aspect of competitive advantage that allows enterprise to lead on its market. According to the research by Aberdeen Group [1], enterprises implementing ERP have enjoyed an average of 12% operational cost reduction, 13% administrative cost reduction, 10% inventory reduction, and 14% complete and on-time delivery improvement. ERP systems help enterprises to integrate processes and function across the enterprise, so it will be easier to manage the operations of the enterprise. Hence, the management can spend more time in strategic thinking and building relationships with partners and customers.

Great advantages of ERP system come with corresponding costs. An ERP implementation costs around tens of millions of dollars for a medium-sized companies and US$300-500 millions for a large corporation. [2] Despite the very large amount of investment, it has been reported that many ERP systems fail to meet expected outcomes. Recent report on ERP implementation projects reveals that these projects were, on average, 178% over budget, took 2.5 times as long as intended and delivered only 30% of promised benefit. [3]

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 563–571, 2015.
DOI: 10.1007/978-3-662-46742-8_51

IT managers stated that one of the primary cause of IT project failure, including ERP implementation, is the poor project planning and management. [4] One of the most important aspect in ERP project planning is the selection of the ERP system to be implemented. Enterprise should choose the ERP system which fit the most with their needs. Careless selection of the system might cause difficulties in ERP implementation, such as need of high number of customization, incompatible business functions, delay in schedule, as well as largely increasing cost.

Researchers have done some studies in selecting ERP for enterprise. Some studies determine ERP selection criteria based on literature studies, some others use technology-based approach for the system selection. Current available approaches for ERP selection decision making are the regular qualitative analysis on criterion, using analytic hierarchy process (AHP) [8], or using analytic network process (ANP) [11]. The qualitative analysis is hard to assess and compared with each other, whereas the technology-based approaches are quite complex and hard to be implemented directly in a non-technological-based enterprise. Therefore, a simpler and more generic model for ERP quantitative assessment is needed.

This paper proposes a general assessment model for ERP system selection and implementation. The method used is based on studying and analyzing the component of ERP system, aspects to be considered in ERP selection, and critical success factors of ERP implementation. Each component is evaluated based on the risks that come with them and how largely they would affect the ERP implementation. Then, the quantitative weight for each component is measured and can be used in ERP system selection.

The main objective of this paper is to provide a generic quantitative assessment model for ERP selection and implementation. Hopefully, the assessment model in this paper could be easily implemented in enterprises, and could help enterprises to do proper analysis in order to be able to select appropriate ERP system for their enterprises. Proper selection of ERP system hopefully can reduce the possibility of ERP implementation failures in enterprises.

The paper is organized as follows, in section 2 ERP selection and implementation consideration is discussed and analyzed, ERP quantitative assessment model is developed and introduced in section 3, and finally the model is concluded in section 4.This is a Sub-section

2 ERP Selection and Implementation Considerations

As mentioned above, the number of failures in ERP implementation is quite high. One of the main cause of the failure is the mistake in selecting ERP system to be implemented. Selecting an ERP system involves more than interviewing a few vendors. The right ERP provider must be a long-term partner.[5] It is therefore, information system researchers have come with various approaches for aspects to take into considerations when selecting ERP system to be implemented.

The aspects to be taken into considerations are not just software-related aspects, but also enterprise-related aspects and vendor-related aspects. Gürbüz, et al [6] have categorized ERP selection criteria in 3 categories : software-related, customer-related,

and vendor-related. A great software won't be effectively used and will be a useless spending if it is not answering the needs and suitable for the people of the enterprise. And also, vendor's reliability and support is highly needed in the ERP implementation. So, we need to include those two aspects in the consideration criterias.

There are quite a lot of researches that have defined ERP selection criteria. Shown in "Table. 1" below is the aspects to take into consideration in ERP selection from past studies.

Table 1. ERP Selection Consideration

Aspects	*Citations*
Software-related criteria	
Functionality	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012; Wei et al, 2005; Verville et al, 2002; Percin, 2008; Palasinamy et al, 2010
Technical aspects	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012; Verville et al, 2002; Palasinamy et al, 2010
Cost	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012; Wei et al, 2005; Percin, 2008; Palasinamy et al, 2010
System reliability	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012
Compatibility	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012
Implementation time	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012; Wei et al, 2005; Bernroider et al, 2001; Percin, 2008
Enterprise-related criteria	
Ease of customization	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012; Wei et al, 2005; Palasinamy et al, 2010
Better fit with organizational structure	Gürbüz et al, 2012; Baki et al, 2005; Palasinamy et al, 2010
Cross module integration	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012; Wei et al, 2005
Vendor-related criteria	
Support and service	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012; Wei et al, 2005; Bernroider et al, 2001; Percin, 2008; Palasinamy et al, 2010
Domain knowledge	Gürbüz et al, 2012; Baki et al, 2005; van der Vorst, 2012
Reputation	Gürbüz et al, 2012; Baki et al, 2005; Wei et al, 2005

Another important aspect in assessing ERP system is the implementation of it. ERP implementation can be assessed by its critical success factors, which can be understood as the few key areas where 'things must go right' for the implementation to be successful. [13]

ERP implementation requires an ensemble view. It is therefore critical success factors for ERP implementation includes not only technical issues but also contextual

issues including social and cultural impact on the interaction between people and the ERP technology. [13] Shown in "Table. 2" are the aspects needed to be considered in an ERP implementation.

Table 2. ERP Implementation Consideration

Aspects	*Citations*
Technical aspects	
IT infrastructure	Zhang et al.,2002; Upadhyay and Dan,2008; Ibrahim et al.,2008; Abdelghaffar et al., 2010
Technical development – modifications, interfaces, and data conversion	Nah et al, 2001; Ehie and Madsen, 2005
IT Architecture	Remus, 2006
People or organizational aspects	
Top management support/involvement	Zhang et al.,2002; Nah et al.,2003; Ibrahim et al.,2008; Abdelghaffar et al., 2010; Upadhyay and Dan,2008
Business Process Re-engineering	Zhang et al.,2002; Iftikhar and Hassan,2008; Nah et al.,2003
Involvement of organizational members/users	Zhang et al., 2002; Ibrahim et al., 2008
Project Management	Zhang et al., 2002; Ibrahim et al., 2008; Nah and Lau,2001
Change management strategy and programs	Brown and Vessey, 2003; Wood,2010; Kumar and Hillegersberg, 2000.

3 Quantitative Assessment Model

The quantitative assessment model is developed based on the ERP selection considerations and ERP implementation considerations. Criteria in ERP selection covers the aspects need to be assessed prior to choosing the ERP system, whereas criteria in ERP implementation covers the aspects need to be assessed prior to implementing the system. If the ERP system scores well in both criterias, we can highly expect the ERP implementation will be a success.

Because the main purpose of this paper is to help enterprises in selecting appropriate ERP system, the assessment model focuses on the ERP selection criteria. Moreover, improper package selection is also stated to be one of the main reason of ERP failure [29]. Therefore, the assessment aspects are focused on the selection considerations. Aspects in ERP implementation consideration will be the supporting aspects for this assessment model. The table below points the aspects in ERP assessment model and their weight in the quantitative assessment model.

Table 3. Quantitative Assessment Model

Aspect	Means	Weight
Software-related aspects		
Functionality	ERP features and modules Multi-language, and multi-site Response time Permission management Database protection Software flexibility	20%
Technical aspects	System reliability System compatibility System architecture System infrastructure	20%
Cost	Cost and benefits Affordability	5%
Enterprise-related aspects		
Ease of customization	User friendliness Generic tools Availability of tools and utilities Documentation standard Flexibility	10%
Fit with organizational structure	User friendliness Business process fit Organization structure fit Human resource fit	15%
Cross-module integration	Stable module integration within system Compatibility with other system Easy integration across enterprise Platform independence	15%
Vendor-related aspects		
Support and service	Warranty details Guidebook Training lessons 24 hour service consultancy Customer support and service	7%
Domain knowledge	Knowledge of industry Number of experienced consultant Number of client	3%
Implementation-related aspects		
Project management	Top management support Project costs Implementation time User involvement	3%
Change management	Introduction and training Business transition User acceptance test	2%

The quantitative assessment model consists of ten aspects to be evaluated, categorized into four categories : software-related aspects, enterprise-related aspects, vendor-related aspects, and implementation-related aspects. These aspects are taken from

ERP selection and implementation consideration above. Some aspects from above are combined with another aspect to build a simpler assessment model. The weighting of those aspects is based on how crucial those aspects are in affecting the success of ERP system implementation.

Functionality and technical aspects have the highest weight (20%). Functionality is considered as the most important evaluation factor. [5] The main reason of enterprise implementing ERP system is to make us of its functionality in order to easily manage the enterprise. Therefore, functionality has to be considered first in assessing an ERP system. On the other hand, technical aspect is also very important. The choice of hardware and software has a major bearing on the acceptance of a system. [27] According to the Kumar et al. [28], system reliability, which is included in technical aspects, is the second important selection criterion. Incorporating the best business practices of every area as well as the latest trends in IT are important for the new system.

Fit with organizational structure and cross-module integration have the second highest weight (15%). According to Ghosh [29], one of the main reasons in ERP implementation failure is the incompatibility with enterprise's business processes. It can be described as gaps between software functions and organizational requirements. Incompatibility might cause the implemented system can't support the enterprise's business and might lead to the huge needs of software customization. Cross-module integration affect the stability and the use of the system. Shikarpur [30] stated that the real benefit of an ERP system is in integration. If the modules in the system and across the enterprise are not compatible with each other, it might obstruct the enterprise integration, which eliminates the main purpose of the ERP system.

Ease of customization is next on the weighting list (10%). Because of the uniqueness in all enterprise, most firms need to customize a part of the ERP system [5]. Customization could bring disaster to ERP implementation if it is too much and too complicated. More experienced consultants and developers might be needed and cause the implementation time and cost to soar up, exceeding the budget and schedule. It would be better if ERP vendor support this, by having proper product documentation, providing high system flexibility, or giving out tools for customization. So, the customization can be easier and cheaper.

Support and service from the vendor is also a quite important aspect. Because installation and ongoing costs can reach seven to ten times the initial software cost, the service and support associated with the application becomes vital to the success of the partnership between end user and application vendor. [5] Following the case that there might be need for technical support in the ERP implementation and maintenance, the presence of IT expertise, in this case is vendor, might be needed. Domain of knowledge of the vendor is also need to be evaluated. Sufficient knowledge of the industry might increase the product compatibility and ERP problem solving capability of the vendor, thus can provide better service to the enterprise.

Following the huge advantage of ERP system, the cost of it is also very high. Therefore, affordability is an important criterion in selecting process; the solution should have attractive prices [31]. Setting realistic expectations for the overall cost of the system is essential to gain top management support, which highly affect the success of ERP implementation.

Implementation-related aspects hold a total of 5% in the aspects weighting. It is a bit low because it has more to do with people/organizational aspect, rather than the

ERP system itself. But, it can also evaluate the ERP system. The project management can be easier if the system has high functionality, fit with organizational structure, and also affordable, which can result in total top management support and shorter implementation time. Change management can also be easier if the program is compatible, stable, and easy to use. So, these aspects can be used in assessing an ERP system.

4 Conclusion

ERP system is a very powerful for managing enterprise. However, a lot of ERP implementation system failed, and one of the main reason is because of improper ERP package selection. Researchers have come with various ERP implementation criteria, but most of them only examine the criterion qualitatively, making it is still hard to choose a better and more suitable ERP system. Therefore, this paper provide the quantitative assessment model for ERP selection and implementation.

The quantitative assessment model is developed based on ERP selection criteria and implementation critical success factors from various citations. The assessment aspects are focused on the selection considerations, supported by aspects in ERP implementation considerations. The weighting for these aspects are determined on how important that aspect is and how crucial these aspects are in affecting ERP implementation success/failures.

The two most important aspects to be considered are the functionality and technical aspects of the ERP system. Other main aspects to be considered are fit with organization structure, cross-module integration, and ease of customization. High scores in these criterions means that the corresponding ERP system is more likely to be suitable for implementation in the enterprise.

Acknowledgement. The research for this paper could be done thanks to the funding "Dana Hibah Penelitian Desentralisasi ITB" from Institut Teknologi Bandung.

References

1. Castellina, N.: A Guide for a Successful ERP Strategy in the Midmarket: Selection, Services, and Integration. The Aberdeen Group
2. Mabert, V.A., Soni, A., Venkataramanan, M.A.: Enterprise Resource Planning: Common Myths Versus Evolving Reality in Business Horizons, May-June 2001, pp. 71–78 (2001)
3. Zhang, Z., Lee, M.K.O., Huang, P., Zhang, L., Huang, X.: A framework of ERP systems implementation success in China: An empirical study. Int. J. Production Economics 98, 56–80 (2005)
4. Umble, E.J., Umble, M.M.: Avoiding ERP Implementation Failure in Industrial Management, vol. 44(1) (2002)
5. Baki, B., C¸akar, K.: Determining the ERP package-selecting criteria. Business Process Management Journal 11(1), 75–86 (2005)
6. Gürbüz, T., Alptekin, S.E., Alptekin, G.I.: An Integrated Decision Support System for Selecting Software Systems. In: eKNOW 2012: The Fourth International Conference on Information, Process, and Knowledge Management (2012)

7. van der Worst, C.: Approach for Selecting ERP Software at Mid-sized Companies Reflecting Critical Success Factors. Journal of US-China Public Administration 9(9), 1057–1068 (2012) ISSN 1548-6591
8. Wei, C.-C., Chien, C.-F., Wang, M.-J.J.: An AHP-based approach to ERP system selection. International Journal Production Economics 96, 47–62 (2005)
9. Bernroider, E., Koch, S.: ERP selection process in midsize and large organizations. Business Process Management Journal 7(3), 251–257 (2001)
10. Verville, J., Harlingten, A.: An investigation of the decision process for selecting an ERP software: The case of ESC. Management Decision 40(3), 206–216 (2002)
11. Percin, S.: Using the ANP approach in selecting and benchmarking ERP systems. Benchmarkin: An International Journal 15(5), 630–649 (2008)
12. Palanisamy, R., Verville, J., Bernadas, C., Taskin, N.: An empirical study on the influences on the acquisition of enterprise software decisions. Journal of Enterprise Information Management 23(5), 610–639 (2010)
13. Xue, Y., Liang, H., Boulton, W.R., Snyder, C.A.: ERP implementation failures in China:Case studies with implications for ERP vendors. In: Int. J. Production Economics, vol. 97, pp. 279–295 (2005)
14. Zhang, Z., Lee, M.K.O., Huang, P., Zhang, L., Huang, X.: A framework of ERP systems implementation success in China: An empirical study. Int. J. Production Economics 98, 56–80 (2005)
15. Upadhyay, P., Dan, P.K.: An explorative study to identify the Critical Success Factors for ERP implementation. In: Indian Small and Medium Scale Enterprises, p. 978. IEEE, 978-0-7695-3513-5/08
16. Ibrahim, A.M., Sharp, J.M., Syntetos, A.A.: Syntetos: A Framework for the implementation of ERP to improve business performance: A case study. In: European and Mediterranean Conference on Information Systems 2008, pp. 25–26 (2008)
17. Abdelghaffar, H., Azim, R.H.A.: Significant Factors Influencing ERP Implementation in Large Organizations: Evidence from Egypt. In: European 2010 Mediterranean & Middle Eastern Conference on Information Systems, Abu Dhabi, Dubai (2010)
18. Nah, F.F., Lau, L., Kuang, J.: Critical Factors for Successful Implementation Enterprise Systems. Business Process Management Journal 7(3), 285–296 (2001)
19. Ehie, I.C., Madsen, M.: Identifying critical issues in enterprise resource planning (ERP) implementation. Computers in Industry 56, 545–557 (2005)
20. Remus, U.: Critical success factors of implementing enterprise portals. Paper presented at 39th Hawaii International Conference on System Sciences (HICSS-39), Kauai, Hawai (2006)
21. Nah, F.F.-H., Zuckweiler, K.M., Lau, J.L.-S.: ERP implementation: chief information officers' perceptions of critical success factors. International Journal of Human–Computer Interaction 16, 5–22 (2003)
22. Syed Iftikhar, H.S., Hassan, S.: Critical Success Factors in ERP Implementation in Pakistan, MS (SE) Thesis, International Islamic University Islamabad (2008)
23. Nah, F.F.-H., Lau, J.L.-S.: Critical success factors for successful implementation of enterprise systems. In: Business Process Management, pp. 285–296 (2001)
24. Brown, I.V., Vessey, I.: Managing the next wave of enterprise systems: leveraging lessons from ERP. MIS Quart. Exec. 2, 65–77 (2003)
25. Wood, B.: SAP ERP Project Failures Lessons Learned and Mini Case Studies 3, in R3NOW (2003), `http://www.r3now.com/tag/erp-criticalsuccess-factor` (accessed September 25, 2014)

26. Kumar, K., Hillegersberg, V.J.: ERP experiences and evolution. Communications of the ACM 43, 23–26 (2000)
27. Poon, P., Wagner, C.: Critical success factors revisited: success and failure cases of information systems for senior executives. Decision Support Systems 30, 393–418 (2001)
28. Kumar, V., Maheshwari, B., Kumar, U.: An investigation of critical management issues in ERP implementation: empirical evidence from Canadian organization's. Technovation 23, 793–807 (2003)
29. Ghosh, R.: A Comprehensive Study on ERP Failures Stressing on Reluctance to Change as a Cause of Failure. Journal of Marketing and Management 3(1), 123–134 (2012)
30. Shikarpur, D.: The dilemma of buying ERP, Dataquest India (October 1997), http://www.dqindia.com/Oct159/3ij1141101.html
31. Rao, S.S.: Enterprise resource planning: business needs and technologies. Industrial Management & Data Systems 100(2), 81–88 (2000)

The Implementation of Customer Relationship Management: Case Study from the Indonesia Retail Industry

Leo Willyanto Santoso, Yusak Kurniawan, and Ibnu Gunawan

Department of Informatics, Petra Christian University
Surabaya 60296, Indonesia
leow@petra.ac.id

Abstract. PD.Cemara Sewu is a company engaged in distributor and selling photography product. Nowadays, system of sales and storage of goods still be done manually. So, company needs a system in order to improve good relationship with customers, and storage goods could be more effective and efficient. In this research, the application process begins with the design to the implementation of the system. Designs are made include the manufacture of DFD and ERD. Application of Electronic Customer Relationship Management that have been implemented to make the buying and storage of goods in PD. Cemara Sewu become more organized, well-structured and have relationship with customers. The application of this system has been created with features that have been planned and tailored to the needs of previously so customer have loyalty to company and reports of sales, card stock and forecasting from the application are accurately, because have the same result with manual calculations. Based on the evaluation, 85% of users said that the features in the application made is considered good and in accordance with company needs.

Keywords: Customer relationship management, customer, implementation.

1 Introduction

In the industrial world, the competition can not be separated from each company. Every company that is engaged in manufacturing or services in general, aims to get the maximum profit and reduce costs. Customer Relationship Management (CRM) is a business strategy that results are expected to optimize the level of profits and customer satisfaction. This is done by way of organizing customer segments, directing the organization so oriented to customer satisfaction and implement business processes that focus on the customer.

CRM is still relatively new in the world of marketing. CRM in the late 90s are felt discussed in many media. It's just that his approach is not much reviewed by the marketers. CRM is a key concept in managing the company's relationship with its customers to create value-added (Value Creation) for its customers. The goal is not to maximize the sales of a transaction, but rather to build ongoing relationships (continued) with

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 572–578, 2015.
DOI: 10.1007/978-3-662-46742-8_52

its customers. Both buyers and sellers are willing to build ongoing relationships during the relationship is a reciprocal relationship that adds value for both parties.

Excellence competition is not solely based on the quality of the product, or the price, but also on the ability of the company to help customers create and develop value-added to them. So as to create customer loyalty, causing the customer will constantly consume products that are created.

Customer loyalty should be improved continuously, otherwise it could be that customers will use similar products of other companies. There is an expression that says, "the customer is king". The phrase fits perfectly with the concept of CRM is aimed at customer satisfaction. If customers are satisfied then it will create customer loyalty. In the business world customer satisfaction is very significant for the progress and survival of a company, many companies fail because it can not retain customers.

PD. Cemara Sewu is a company engaged in the distribution and sale of products for photographic purposes. A company based in Jakarta has several branches, among others, Surabaya, Semarang, Bandung and others. In the PD activity. Cemara Sewu sought in order to satisfy the customers so as to get the maximum benefit. We need a new model in Customer Relationship Management to manage the relationship between the distributor and the customer as well as the procurement of products that are more efficient and effective in its efforts to increase profits by reducing costs of logistics activities.

The purpose of this paper is to increase the income of the customer satisfaction with the use of electronic customer relationship management (e-CRM) in PD. Cemara Sewu.

The paper is organized as follows: Section 2 reviews the literature on CRM implementation. In Section 3 we have presented the CRM implementation in PD Cemara Sewu. Section 4 discuses the implementation and analysis of CRM implementation in PD Cemara Sewu. Finally Section 5 draws conclusions from the case study in terms of its practical relevance and lessons learned

2 CRM

Customer Relationship Management is used to define the process of creating and maintaining a relationship with a customer-customer business or customers. CRM is the process of identifying, attracting, and retaining customers differentiate [5]. According to Craig Conway, CRM is the ability to recognize the transaction experiences faced by customers for transactions with companies in which the CRM try to improve customer satisfaction to loyalty and tendency of customers to buy increasing [3]. CRM applications can allow management or setting a good relationship with customers effectively, supported by the availability of good leadership [3].

The purpose of the CRM business framework is as follows [4]:

1. Adding existing relationships to increase revenue. Enterprise wide views of customers to maximize the relationship between them so as to improve profitability companies to identify, attract, and retain potential customers.
2. Using the integrated information for the best service. By using customer information to provide better service for their needs, so customers do not have to repeatedly ask for the information they need to companies that save time and reduce their frustration.

3. Introducing channel processes and procedures those are consistent and replicable. With the development of communication channels for customers, the more employees engaged in sales transactions, so companies need to improve the consistency of the process.

So the goal of CRM is to have a relationship with customers that can provide significant benefits to the company. To achieve these objectives, marketing, sales and customer service should work more closely in teamwork and sharing of information. CRM consists of the following phases [4]:
1. Getting new customers (Acquire)
2. Increasing customer value (Enhance)
3. Maintain existing customers (Retain)

Some types of CRM technologies include [6]:
1. Operational CRM
   Integrated automation of business processes, including customer touch points, channels, frontback office, integration
2. Analytical CRM.
   Analysis of the data generated by the CRM operations, including data mining applications.
3. Collaborative CRM.
   Collaboration applications including e-mail, personalized publishing, e-communities, and the like are designed for interaction between the customer and the organization.

2.1 Customer Profitability Analysis

From company's perspectives, the customer plays a very important role in addition to the growing importance of customer positions, companies face the fact that its resources are limited. It requires that companies are increasingly selective and effective in allocating these resources to customers who can provide benefits to the company. Customer profitability analysis is an approach to cost management that identifies the cost and benefit of serving specific customer or customer type to improve overall profitability an organization [1].

Customer Profitability Analysis has two main objectives, namely:
1. To measure the profitability of existing customer. Customer profitability analysis can show the cost benefit analysis to identify profitable customers or not.
2. To effectively identifying whether or not consumer-related activities. This analysis provides information that can be used by organizations to decide which activities need to be maintained or reduced so as to improve profitability.

There are several stages in conducting Customer Profitability Analysis, namely:
1. Identifying customer
2. Calculate the cost of customer revenue and customer
3. Analyze where customers are profitable and which are not.

3 Design and Analysis of System

Based on the analysis of company's old system, there are a number of problems. It needs to solve the problems found in the old system, with the new system. The new system requirements are:

1. An integrated system for recording sales, procurement, stock, and all sorts of business activities.
2. Customer can see the information section of the company's warehouse and the amount of the price, so that customers can make their own, and direct sales order will be sent to the company and order status can be monitored.
3. Customer can do setting of the minimum stock of an item that is often ordered to the company so that if the customer has a minimum stock in the order will be immediately made.
4. Company and customer can do sales forecasting.

Design of the CRM at PD Cemara Sewu are as follows:

1. Getting new customers (Acquire)
   Getting new customers by:
   - Promoting the products produced by the company.
   - Provide comfort to the customers in buying the products they need. The goal is to offer a good product with satisfactory service.
   - Featured in the program are given a page gallery to display the latest photos of the products offered. And customers can create their own sales order via the web.
2. Increasing customer value (Enhance)
   Companies must create a close relationship with customers by the company to listen to complaints and improve services. Relationship with customers can be enhanced by:
   - Provide different prices according to the profitability of each individual customer.
   - There are features of profitability and customer rankings so that each customer gets different treatment according profits earned by the company.
3. Maintain existing customers (Retain)
   - Making time to listen to the needs of customers, including customer dissatisfaction towards the product or service company. So it can be used to improve services.
   - There is a communication feature (both complaints and promotion) between the customer and company.
   - There are a feature campaign and also alerts for purchasing patterns of each customer so that sales can follow-up customer orders that have exceeded their limits.
   - There is also a feature of forecasting for the sale of each item.

The system is designed to involve three external entity, PD Cemara Sewu main office, customers, and branches of PD Cemara Sewu. Customer is the entity that is

engaged in transactions with the company in sales activities ranging from sales order, the minimum stock, and others. PD Cemara Sewu main office is an entity that provides goods to be sold by the company in purchasing activities include the purchase invoice, and others. While branches of PD Cemara Sewu is the entity that receives financial statements and reports required by the company forecasting to determine the direction and policies of the company. Fig. 1 is a diagram of the design context of customer relationship management.

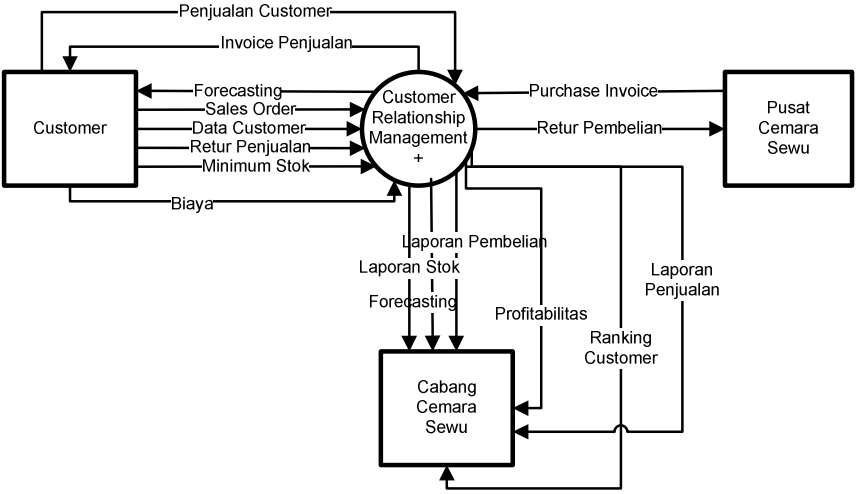


Fig. 1. The Context Diagram of the Application

The implementation primarily consisted of the major steps as given in Table 1.

Table 1. Major tasks during implementation process

No	Tasks	Duration
1	**Infrastructure readiness**	**3 days**
	Server setting and installing	1 day
	PC and Notebook connecting	2 days
2	**Mapping process**	**5 days**
	Data gathering	3 days
	Database process mapping	2 days
3	**Data cleaning**	**15 days**
4	**User training**	**5 days**

4 Analysis

This chapter discusses the testing system has been created. The test system is implemented as a whole one by one according to the feature of the application.

This login page is the home page at the time the program is run. To be able to run the program/operate the program, users are required to login first to enter a username and password that was created earlier. The login page can be seen in Fig 2.

Fig. 2. Login Page

The customer profitability reports are used to see the profit from each customer so that the customer can be grouped where customers are making a profit and which are not. Reports for customer profitability can be seen in Fig. 3 and Fig. 4.

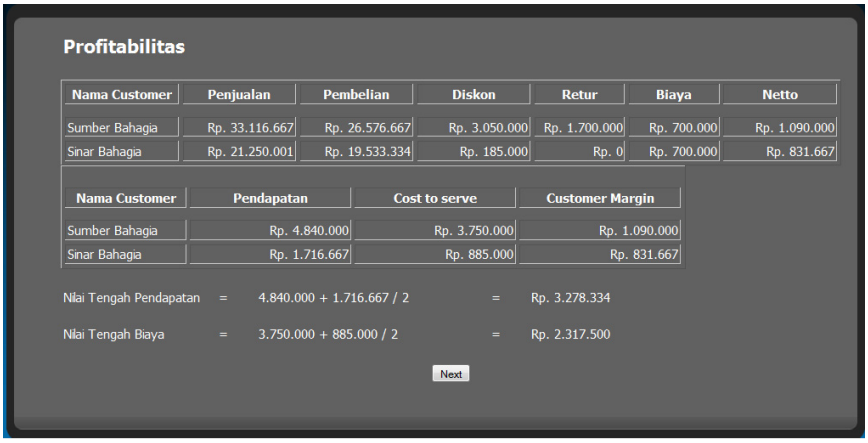
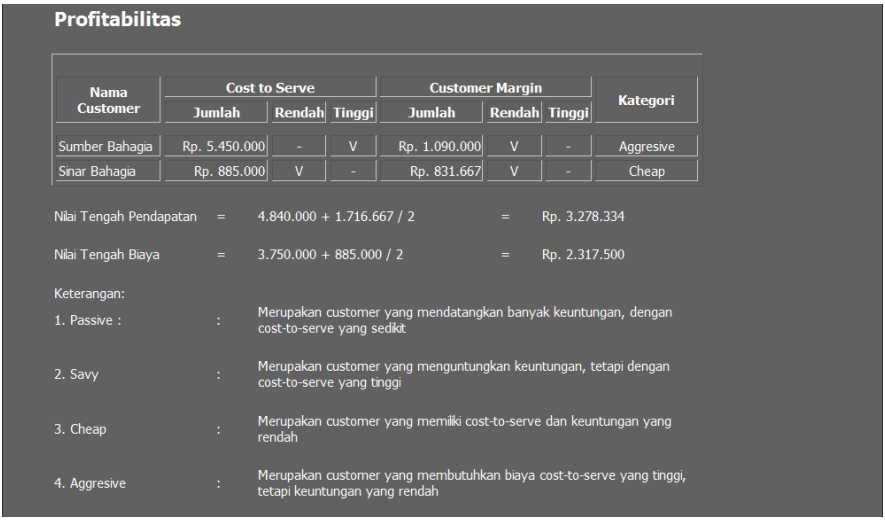


Fig. 3. Customer Profitability Report Page



Profitabilitas

Nama Customer	Cost to Serve			Customer Margin			Kategori
	Jumlah	Rendah	Tinggi	Jumlah	Rendah	Tinggi	
Sumber Bahagia	Rp. 5.450.000	-	V	Rp. 1.090.000	V	-	Aggresive
Sinar Bahagia	Rp. 885.000	V	-	Rp. 831.667	V	-	Cheap

Nilai Tengah Pendapatan = 4.840.000 + 1.716.667 / 2 = Rp. 3.278.334

Nilai Tengah Biaya = 3.750.000 + 885.000 / 2 = Rp. 2.317.500

Keterangan:

1. Passive : Merupakan customer yang mendatangkan banyak keuntungan, dengan cost-to-serve yang sedikit
2. Savy : Merupakan customer yang menguntungkan keuntungan, tetapi dengan cost-to-serve yang tinggi
3. Cheap : Merupakan customer yang memiliki cost-to-serve dan keuntungan yang rendah
4. Aggresive : Merupakan customer yang membutuhkan biaya cost-to-serve yang tinggi, tetapi keuntungan yang rendah

Fig. 4. Customer Profitability Report Page

5 Conclusion

From the results of implementation of CRM in PD. Cemara Sewu, some conclusions can be drawn, namely:

a. The application is considered to be able to answer the needs of the company. The relationship between the company and the customer can be established well.
b. The forecasting allows the company to forecast sales in the following months so out of stock could be eliminated in the company.
c. Based on the user questionnaire, it can be concluded that the system is able to be used properly and assist in the execution of daily business processes.

References

1. Bochler. Introduction to Cost Management. Mc. Graw Hill, New York (2006)
2. Cokins, G.: Activity based cost management, an executive guide. John Wiley & Sons, Canada (2001)
3. Greenberg, P.: CRM at the speed of light: capturing and keeping customers in internet real time, 2nd edn. McGraw-Hill Book, California (2002)
4. Kalakota, R., Robinson, M.: E-business 2.0 roadmap for success. Addison-Wesley, USA (2001)
5. Strauss, J., Frost, R.: E-marketing. Prentice Hall, Upper Saddle River (2001)
6. Turban, E., Rainer Jr., R.K., Porter, R.E.: Introduction to information technology. John Wiley & Sons, USA (2005)

The Implementation of Customer Relationship Management and Its Impact on Customer Satisfaction, Case Study on General Trading and Contractor Company

Djoni Haryadi Setiabudi, Vennytha Lengkong, and Silvia Rostianingsih

Informatics Department, Petra Christian University
{djonihs,silvia}@petra.ac.id

Abstract. This study concern about the implementation of Customer Relationship Management (CRM) and the impact on customers satisfaction of General Trading and Contractor company from various industrial sector. The target of the company is increasing the profit by handling personally every customer or future customer. However, the fact showed that the manager found difficulties in controlling the performance and progress of his sales. This condition makes the service schedule which is promised to the customer sometimes cannot be fulfilled on time. This problem happens because there is no reminder system to the manager about the service schedule. The implementation of Customer Relationship Management in this company includes: make the assignment for sales in handling the customer or future customer, reminder, online questionnaire, complaint, added wish list, and system to show notification. The result of this study shows that the customer satisfaction can be increased as much as 92%, with the implementation of the following features: system can create marketing campaign, customers can make a complaint online, employee can create reminders to remind the schedule of services and other purposes.

Keywords: Customer Relationship Management, satisfaction, service, reminder, online questionnaire, complain, wish list.

1 Introduction

This research focus on the implementation of Customer Relationship Management (CRM) in CV X and its impact on the satisfaction to its customers. CV X is a General Trading and Contractor company handles sales, rental, and service of power plants devices beside of canopy and lifts.

Currently CV. X using spreadsheets in recording the sale or rental from customers. Although it has been used spreadsheet, CV. X still has difficulties in checking all the sales that come from customers. Another problem faced by CV. X is in terms service. CV. X often forget to schedule services that have been made since it only recorded manually, so CV X need an online reminder system to alert an appointment has been made.

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 579–587, 2015.
DOI: 10.1007/978-3-662-46742-8_53

While the service is done in case of product damage, usually delivered in the form of a complaint. But the complain that do not require repair are usually not recorded. Owner wants to know any complaints from customers in order to make improvements both in terms of services and products sold or leased. In order that all complaints will be managed properly, it is necessary that the complaint features can be accessed by customers directly via website.

Customer ratings are important in business continuity. CV. X uses a feedback questionnaire to measure customer satisfaction after using the products or services provided. The questionnaires were distributed by sales to customers. Owner wants to make sure that the sales do not cheat in terms of filling out the questionnaire. To answer these doubts it needed an online questionnaire so that customers can directly provide the answer.

In addition to the questionnaire, the owner also wants to control the performance of the sales, so that the operating costs are proportional with the resulting performance. Sales role is very important because it deals with customers or prospective customers directly, so it needs a system that can control the performance of the sales in prospecting the prospective customers and existing customers.

Some researchers have been conducted a study on implementation of CRM on companies. Mithas et.all [1] state that using of CRM applications is positively associated with improved customer knowledge and improved customer satisfaction. Lawson-Body and Limayem[2] made the study to explain the impact of Web site characteristics on the relation between customer relationship management (CRM) and customer loyalty. They found that using the Internet to support CRM allows firms to increase their customer loyalty in the IT sector.

Becker et.al. [3] conduct the study to analyze what influence companies can expect CRM implementation to have on performance and how they can leverage its impact. The authors propose a conceptual model that investigates the link between technological and organizational implementations, as well as the implementations' interactions with management and employee support and CRM process-related performance.

Obeng and Loria [4] studied of implementing CRM on two Swedish service farm. The study show that services company enjoy some benefits by implementing CRM.

This study develop a web based application to control sales performance, handling the services and complain, marketing campaign, make the assignment for sales in handling the customer or future customer, reminder, online questionnaire, complaint, added wish list, and system to show notification. Finally, it also monitoring the impact on customer satisfaction.

2 Customer Relationship Management

According to Chaffey [5] Customer Relationship Magement is an approach to building and sustaining long-term business with customers. In many literatures the terms of customer relationship management and relationship marketing are used interchangeably. CRM comprises of four marketing activities: customer selection, customer acquisition, customer retention, customer extension.

According to Kotler and Keller [6], Customer Relationship Management (CRM) is the process of managing detailed information about each customer and carefully manage all "touch points" of customers in order to maximize customer loyalty, however Gordon [7] state : “CRM is a series of strategies and processes that create new and mutual value for individual customers, build preference for their organizations and improves business result over a lifetime of association with their Customers”.

According to Kalakota [8], the objectives of CRM are : Using relationships with customers to increase corporate profits, Using information to provide better services, supports the process of repeat sales to customers.

3 Research Methodology

3.1 Analysis

CV. X has customers from various companies. CV X assigning sales to introduce CV X on customers and promote the products. Moreover CV.X also promotes itself through brochure and online media. Not only sales that come to customers and prospective customers, but customers can also contact the CV. X directly if interested in the ads that are posted. The products offered can be either new, complementary products or recommend a product that has more features than the product that has been purchased previously.

In terms of documentation, CV. X just use spreadsheet to record the transactions. Other related transactions such as offers, complaints, and services has not been recording.

Concluded, some of the problems encountered are:

- Difficulty in finding a sales transaction in case of any complaint and service because the documentation is still using spreadsheet.
- Time delayed in periodically service because the recording is still manually so that no alert or notification as a reminder.
- Collection of questionnaires from customers are carried out by sales, so the answer of the questionnaire could have been manipulated by the sales.
- Difficulty in controlling the sales performance in dealing with customers or potential customers. For example, controlling the activities of marketing campaign.

Based on these problems, it can be concluded that the system required by the CV. X is:

- Creating a system for marketing campaigns, catalogs can be offer, sending mass e-mails, assignment to sales and adding product to wish list for customers who require a particular product.
- Create a system for recording each transaction in order to search the data easily and preparing the reports to owner.
- Create a reminder that can be accessed by employees of CV. X to make the scheduling for service and other activities that can be done on time as scheduled.
- Create an online questionnaire to maintain the accuracy of the answers provided by the customer. It also reduces operational costs of CV. X.

The illustration of the required system can be seen in Fig. 1. This illustration shows that the company can send a bulk e-mails to customers to offer products. The next scene shows the customer can add product to wishlist. The other scene shows that customer can send complaint to the system and obtain services.

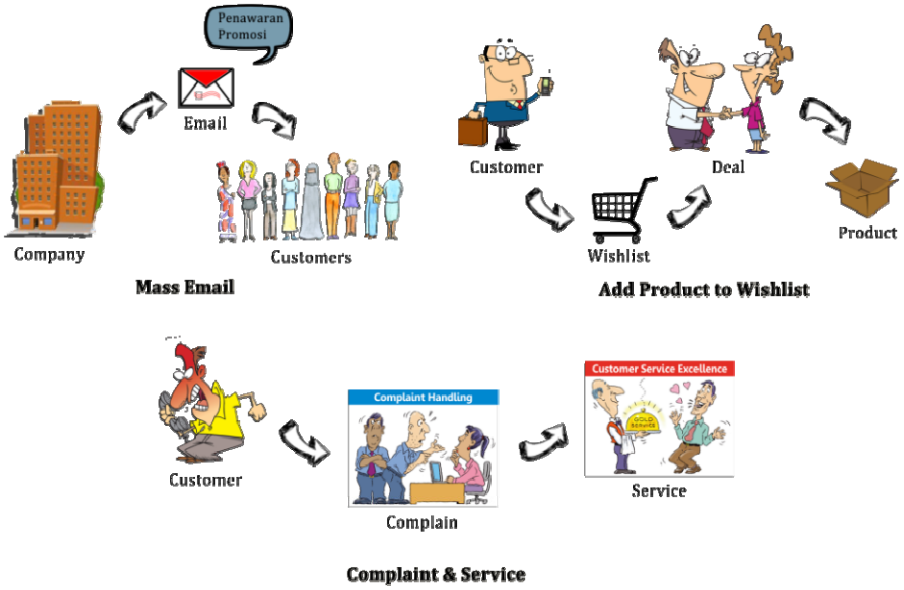


Fig. 1. The Illustration of The Required System

3.2 Application Design

The Data Flow Diagram explain the process of the entry and exit of the data flow from the system. DFD level 0 is shown in Fig 2.

Some important processes are :

- Boards.This part is the assignment process. Manager as a supervisor added a member as the assigned employee. In this section also carried out recording of other transactions, namely offering product, sales, rental, service, and marketing campaign.
- CRM. This section is a process for dealing with a service that can be accessed by customers / potential customers, namely complaints, catalogs, wish list, and questionnaires.
- Features.This section is a process in the company used for activities related to the customer. There are reminders and email.
- Report.This process creates reports of the transactions dan income to owner.
- Notifications. This process will provide notification to other users who are concerned with these activities. Notifications occur for several processes, namely:

- When an employee is added as a member on a board or card.
- During the activities associated with the card.
- When customers fill out a questionnaire.
- When customers add wishlist.
- When adding a customer complaint.
- When customers or potential customers register to receive access rights on the website.

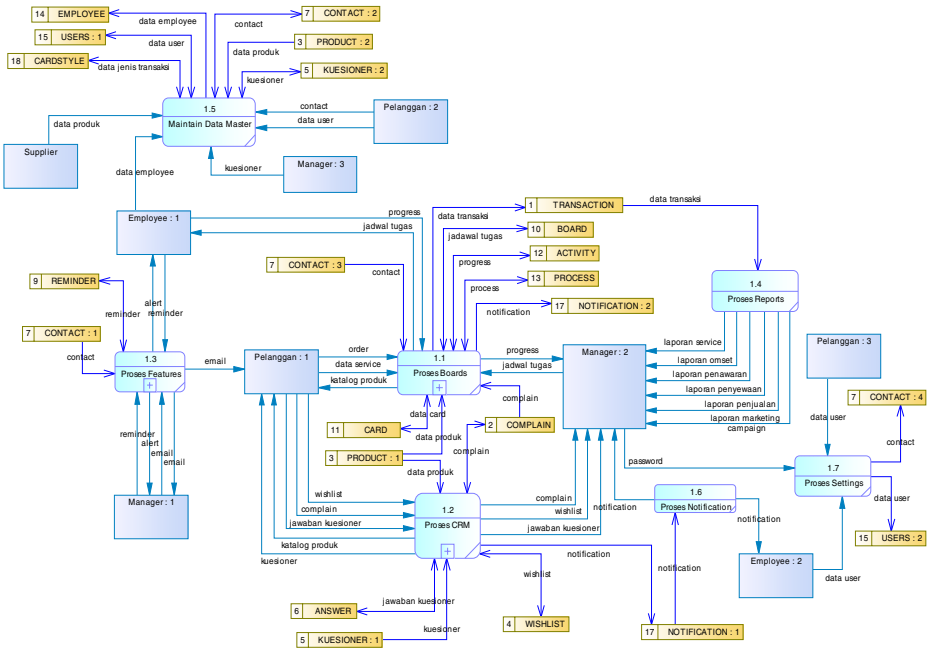


Fig. 2. Data Flow Diagram of The Application

4 Experimental Results and Case Studies

Boards

The menu boards is a menu used to make assignments to sales in dealing with customers / potential customers.

Case Study: DN as a sales is assigned by M as sales manager to handle PT BJA as a customer in generator sales. Manager wants to control every progress of DN. PT. BJA give a positive response to using the product of CV. X. After finding a deal price, then PT. BJA asked to give down payment. Furthermore DN is assigned to handle the orders of PT. BJA until the goods are received by PT. BJA. Board data entry can be seen in Fig 3.

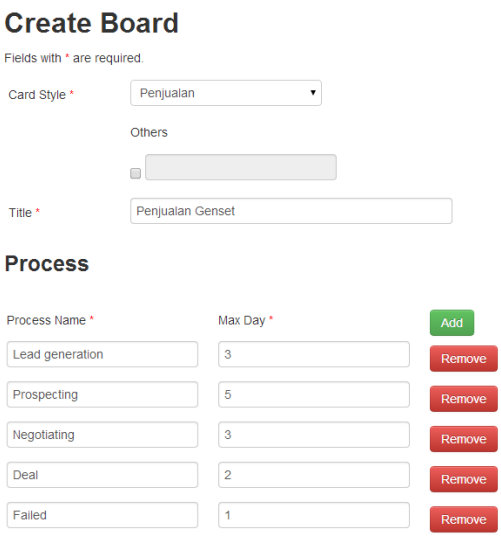


Fig. 3. Create Board Entry

From these assignment, the first data board filled is the card style. Manager wants to control the progress of sales in selling the generators, then the card style selected for this assignment is selling. Processes in this board are "Lead generation, prospecting, negotiating, deal and failed". View Board Page can be seen in Fig 4.

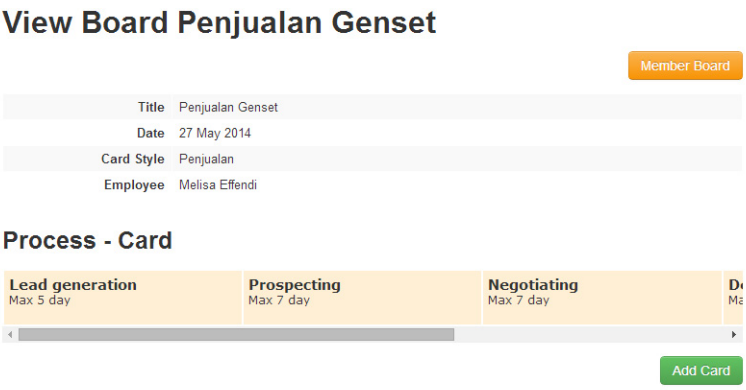


Fig. 4. View Board Page

Marketing Campaign

Case study: CV. X conducting a marketing campaign to attract customers / potential customers to use the products or services offered. M as the manager determines the target customers those are S hotel and PT. I. For the both prospecting, M assigning V and S to deal with each of these prospects. The process is focused on marketing campaign of the lift product. Sales then contact prospects by sending each packet such as vouchers or flyers to the prospect. Sales then confirm again to determine whether the

packet was received or not. The approach followed to determine the needs of the target. Hotel S and PT S has a tendency to use the lifts product of CV.X. Because both targets show a good response, then the marketing campaign for the lift is considered successful. M then assigning its sales further assignment to lift sales. View Board for the assignment of marketing campaign can be seen in Fig. 5.

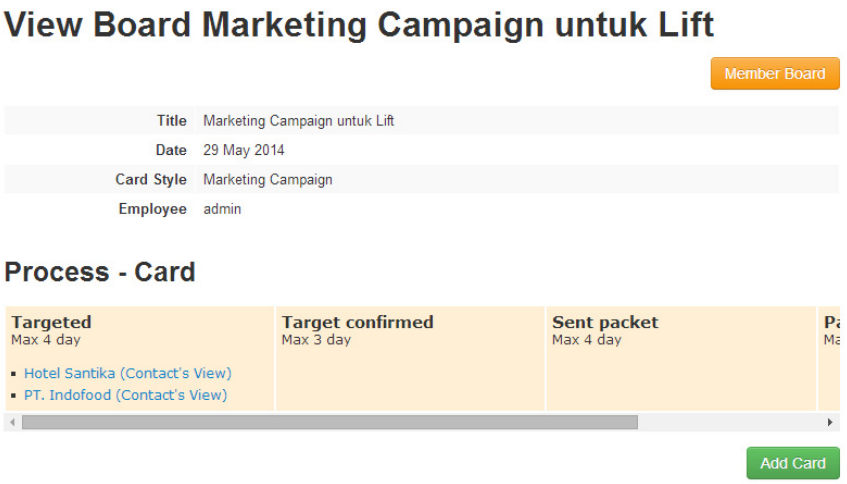


Fig. 5. View Board of Marketing Campaign

Reminder

Reminder used to create a reminder. Reminder index page displays a reminder of data ever made by him (the user who is logged in). Page to create reminders can be seen in Fig. 6. Case study: CV. X has scheduled service for the lift of PT. BJA on May 29, 2014. Manager would like to remind the sales Kusmin three days before the scheduled service so Kusmin can prepare everything necessary for the service.

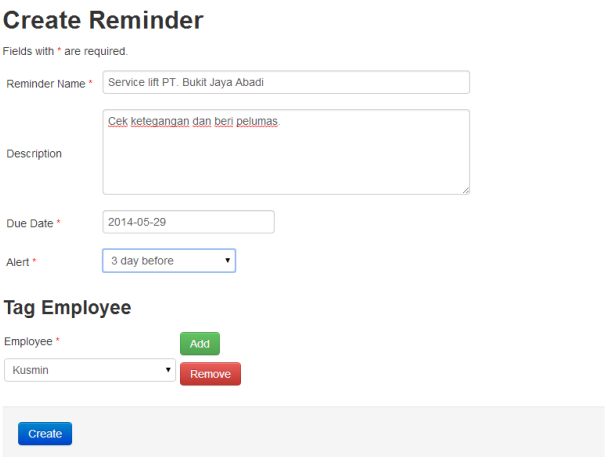


Fig. 6. Create Reminder Page

Complain

Case Study: Customer Mega bought Man Marathon generator 300 KVa silent type one month ago. But a few days later the generator cannot be started. Mega then submit complaints to the CV. X. Mega entered the complain about Man Marathon product with 300 kva silent type. The selected board is "complain generator", then for the description is filled with "generator cannot be started". Add Complain Page can be seen in Fig 7.

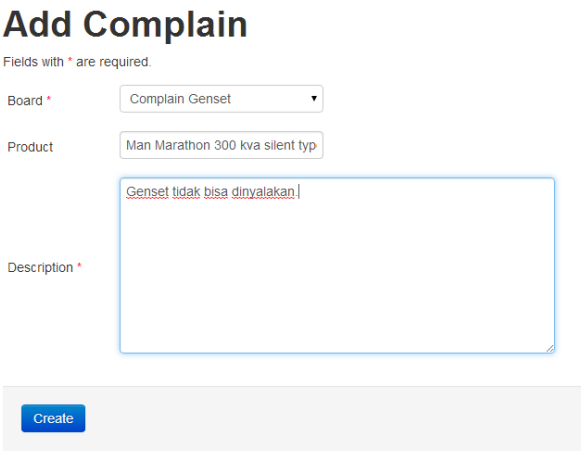


Fig. 7. Add Complain Page

5 Conclusions

- Using this application, managers can control the assignment to sales in dealing with customers / potential customers through the board
- The system can send mass email for various purposes. The system can record the transactions that have been done.
- The system can provide alerts to a user on a specific date based reminders are made.
- Customers can make a complaint online.
- The customer / prospective customers can see the product catalog and make additions wish list.
- The result of this study shows that the customer satisfaction can be increased as much as 92%, with the implementation of this application

References

1. Mithas, S., Krishnan, M.S., Fornell, C.: Why Do Customer Relationship Management Applications Affect Customer Satisfaction? Journal of Marketing 69(4), 201–209 (2005)
2. Lawson-Body, A., Limayem, M.: The Impact of Customer Relationship Management on Customer Loyalty: The Moderating Role of Website Characteristics. Journal of Computer Mediated Communications 9(4) (2004)

3. Becker Jan, U., Greve, G., Albers, S.: The impact of technological and organizational implementation of CRM on customer acquisition, maintenance, and retention. International Journal of Research in Marketing 26(3), 207–215 (2009)
4. Obeng, T.K., Loria, K.: Customer Relationship Management, A case Study of Two Service Companies, Master Thesis, Lulea University of Technology (2005)
5. Chaffey, D.: E-Business and E-Commerce Management, Strategy, Implementation and Practice, 4th edn. Prentice Hall (2009)
6. Kotler, P., Keller, K.L.: Marketing Management, 13th edn. Prentice Hall (2009)
7. Gordon, I.: Best Practices: Customer Relationship Management. Ivey Business Journal (2002)
8. Kalakota, R., Robinson, M.: E-Business 2.0 Roadmap For Success. Addison-Wesley, USA (2009)

Towards e-Healthcare Deployment in Nigeria: The Open Issues

Jumoke Soyemi, Sanjay Misra, and Omoregbe Nicholas

Department of Computer Science
Covenant University, Ota, Nigeria
jumokesoyemi@yahoo.com,
{sanjay.misra,nicholas.omoregbe}@covenantuniversity.edu.ng

Abstract. Information and communication Technology (ICT) has played vital roles in so many disciplines and it is one of the driving forces behind globalization. It has closed the gap in communication as well as enhanced prompt decision making thus accomplishing the United Nation Millennium Development Goals Target 18 which expresses the maximization of the values of advancement in ICT through collaboration with private sectors. Healthcare as a discipline has also embraced this dynamic tool to fashion out what is known as e-health targeted towards improving the health system of the people. Consumers are tired of the usual routine of waiting long hours on queue for appointments; struggling with inconvenient scheduling that deprives them of being fully at work or engaging in productive work. The fact still remains that healthcare system in developing countries has not kept pace with other sectors of the society especially in embracing the power of ICT. This paper reviews the open issues facing the adoption of e-healthcare in Nigeria, highlighting issues that were not discussed in the papers reviewed and giving plausible solutions for its effective adoption and expansion. In addition a model was proposed for e-healthcare adoption in Nigeria.

Keywords: Conventional healthcare, e-Healthcare, Electronic health record, Information and communication Technology, Open issues.

1 Introduction

The rules and roles of the delivery of healthcare continue to change drastically in the 21st century and operators are more interested in cost reduction, greater efficiencies and effectiveness. Qualitative delivery, quick and widespread accessibility are key watchwords. Healthcare is the treatment diagnosis and prevention of illness, diseases and other physical and mental deficiencies in human beings. It is the work done in providing primary care, secondary care, tertiary care and public health. Healthcare system is a process of organizing, integrating, storing and retrieving patient and healthcare information which has been paper based. The paper based method has not been efficient especially for managing the huge amounts of patient and healthcare information as affects patient care. For instance the healthcare records could be illegible

R. Intan et al. (Eds.): ICSIIT 2015, CCIS 516, pp. 588–599, 2015.
DOI: 10.1007/978-3-662-46742-8_54

because it is handwritten and poorly structured, making it difficult to track the information needed on past medical tests and results [1]. There are certain patients that registered with several healthcare providers and these patient records are never shared with other physicians, laboratories and hospitals. Hence information becomes fragmented causing disruption, delay and error in patient care [2].

Patient most times does not have access to their accurate and reliable information which could be used by them to meet their need. Studies revealed that patients who understand their condition and are involved with doctors in making decisions are better able to deal with their illness or diseases [3]. The incidence of September 11, 2001, and their aftermath also put countries like the United States, United Kingdom and so on to be on alert and pushed them to develop additional computer systems that can collect, analyze, coordinate and distribute health information. Much more is the fact that physician and public health workers need to be able to respond quickly and effectually to bioterrorism [4].

This paper reviews open issues identified and those that are not identified in literature hindering Nigeria from successfully transiting from conventional healthcare system to e-healthcare system and also proffer plausible solutions to those issues.

The general introduction was done in section 1, the background of transition from conventional healthcare to e-healthcare in developing countries was given in section 2. We demonstrated the existing e-health systems especially in developed countries and also the benefit of the system in section 3. We have discussed the issues, challenges and solution on e-health specific to Nigeria in section 4, with a table highlighting the open issues identified in literature as well as those not identified giving plausible solutions. A model of healthcare transition in Nigeria was proposed. Finally conclusion was done in section 5.

2 Background of Transition to Electronic Format (e-Health)

Over the decades, a good number of organizations invested greatly in turning away from paper data storage into electronic platform and format, thereby providing a convenient way of managing data. e-health which is a field of medical informatics is the delivery of health services and information through the use of internet and other related technologies. Specifically, e-health can be defined according to Mitchell [5] as the use of ICTs in the health sector for clinical, administrative and educational purposes either at a local site or at a distance thus accomplishing the UN Millennium Development Goals Target 18 which states that "In co-operation with the private sector, make available the benefits of new technologies, especially information and communications technologies" [6]. It is a new way of healthcare delivery different from what used to be [7]. The main application domains of e-health includes: Electronic health records, Decision support tools, Telemedicine and Telecare services. The major goal pursued by e-health policy focuses on the healthcare quality improvement and cost reduction [8], [9]. This goal encapsulates e-health several promises which is reviewed in Lapointe [10].

Some of its objectives are, in-patient cost reduction and improved health outcome; improvement in prescription adherence through reminders and telemonitoring; reduction in

medical errors and drug adverse event using computerized reporting system, e-Prescription of diagnostic procedures, electronic records and so on. e-health promises improvements, accessibility, effectiveness and efficiency of healthcare delivery. This is guaranteed on the basis of compliance with fundamental requirements in term of quality and safety.

The major stakeholders in the e-healthcare industry according to Austin and Boxerman [11] are employers, patients, providers and health plans. The employer wants to analyze healthcare cost and utilization by their employees, the patients want prompt information about their own health while the providers want to save money and time by streamlining communications and health plans strengthen relationships with members while also reducing cost of doing business. Several developed countries such as United States (US), United Kingdom (UK), Canada and others have advanced and passed into law policies, guidelines and regulations aim at enticing providers (Physicians and hospitals) to adopt e-Health. Example of such is Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 in United States [12]. Over time, many e-healthcare applications have been developed for healthcare delivery and a review of four of them is hereby presented.

3 Existing e-healthcare System and Benefits

There are quite a number of existing e-healthcare systems. Few basic ones are reviewed here. They are Computerized Physician Order Entry (CPOE), Computerized Adverse Event Systems, Clinical Decision Support (CDS) and Electronic Health Records (EHR). These four have been chosen because they are the basics for e-healthcare system startup. Others not discussed includes but not limited to Telemedicine, Telecare, Incidence Reporting System and so on. They are not discussed because the feasibility is slim at this early stage of e-healthcare startup in a developing country like Nigeria.

3.1 Electronic Health Records (EHR)

The Electronic Health Records (EHR) holds all the data and medical history, medications, diagnosis, laboratory test results and so on of the patients in an electronic format. The EHR guides against the loss of patients records and put a greater control on medical prescription. This guarantees patients safety and provides real-time firsthand information at the disposal of health professionals for quicker response in the treatment of the patients. One major benefit is the ease of access by authorized personnel and ability to share information across a wide range of healthcare providers.

3.2 Clinical Decision Support (CDS)

A clinical decision support (CDS) helps the operators in arriving at a decision as it concerns their clients (patients). Most CDS functions include offering the current information about a drug, alerts for drug interactions and checking a patient allergy to drugs. Each function provides another means of caring for the patients in a safer and efficient ways. All in all, the patient will be better off with more safer and effective care as the CDS is used and to a large extent, medical errors will be averted [13][14].

3.3 Computerized Physician Order Entry (CPOE)

The CPOE is a process where the physicians' order regarding the patients' treatment under their care are inserted electronically and transferred directly to trusted and reliable services or individuals. What was obtainable in the past were orders written by hand or verbally conveyed to the patients which sometimes leads to medical errors [15], [16]. Virtually all CPOE systems have in-built clinical decision support systems that offer simple and elementary computerized guidance as it affects drug doses and frequencies and to more complex information such as allergy to drugs, drug-laboratory values, drug-drug interactions, guidelines and checks [17]. CPOE are applicable in almost all environments and platforms that currently use available software though it can be time consuming, resource demanding and costly [18]. The deployment of a good CPOE is to make the following attainable: accessibility at anytime and anywhere, reduction in transcription and misconception errors by eradicating medication orders written with hands, reduction in call-backs to physicians for interpretation of unclear, vague, confusing and inadequate order, and time used for prescription and administration of drugs [19].

3.4 Computerized Adverse Event Systems

The computerized adverse event systems is used to observe the rate at which adverse events occur and to alert operators when definite pointers are present which makes it different from CPOE that target to prevent errors. Infections like nosocomial and adverse drug effects events are common adverse events that ICT systems have been used to test effectively [20]. The method employed in most health facilities for reporting of incidences are spontaneous because they rely mainly on reports from physicians, pharmacists and nurses who have direct contact with the patients, to discover drug adverse events which are basically ineffective as it detects about 5% events [20]. One of the outcomes of most trials using computerized adverse event systems is the considerable surge in the number of drug adverse events that were observed and reported which can propel an automatic alert that reduces time for patients' treatments and improve critical laboratory results [21]. Certain adverse events can be detected with tools such as event monitoring, which work well with some adverse events with nosocomial infections and drug adverse events inclusive. This has been well adopted by many health facilities as it detects on time many adverse events.

3.5 Benefits of e-Healthcare

e-health is designed to improve different sectors of healthcare in term of quality healthcare delivery, cost effectiveness, termination of the queuing process to see doctors and lots more. This will be achieved through; improved, accurate diagnosis and appropriate treatment, supportive delivery of care targeted to individual patient based on evidence and patient specific data.

e-health also supports evidence based practice and error reduction, reducing barriers created by physical location and disabilities by improving effective healthcare access. Improving diagnostic accuracy, treatment appropriateness, cost efficiency by

streamlining processes, effective patient administrative systems, reducing waiting time and waste and also empowering patient for self-care and decision making in term of their health. Nevertheless, as e-health matures, it is clear that this is just one side to the added value [22]. E-health has a great potential to empower citizens, patients, and healthcare professionals. It can also offer governments a way to cope with increasing demand for healthcare services, and reshape the expectations of health care delivery, making it people-centered.

4 Issues, Challenges and Solution on e-Healthcare Specific to Nigeria

There are certain issues and challenges peculiar to Nigeria which could hinder e-healthcare deployment from materializing. These issues are discussed and the suggested solutions given below.

4.1 Issues and Challenges

The healthcare sector in Nigeria has not kept pace with the monumental benefits experienced by other sectors of society in embracing information and communication technology [23]. Being one of the most populous black nation with a population of over 160 million people, a projections from Nigeria Censor of 2006 [24] is reported to have one of the highest death rates in Africa, because of inefficient healthcare delivery scheme [25]. With inefficiency in healthcare scheme, Nigeria is likely to be at health risk [26]. The government therefore needs to exploit the potential of ICT like other developed nations of the world, to improve the health sector of the country. [27].

The general open issues facing e-health as it relates to Nigeria includes but not limited to the following:

Electric Power Supply. Nigeria up till now has not succeeded in the provision of uninterrupted power supply to its citizens with a total of less than 4,000MW of power produced [28][29]. ICT services in which e-healthcare is one can only thrive on electric power supply. This is a major issue in the country affecting other sectors aside from ICT. The internet facilities in particular have suffered regular downtime as a result of equipment damage due to interruptions of power. Ghana which most Nigerians likened to be like a state in Nigeria celebrated 10th year of uninterrupted power supply few years back while Nigeria which is seen by most African countries as the giant of Africa is still struggling hard in this regard. For the successful deployment of e-health in Nigeria, the issue of electric power supply must be resolved and the government has a major role to play in achieving the goal. In the absence of electric power supply, alternative power supply (solar, inverters and so on) should be provided.

Internet Facilities. Notwithstanding the successes recorded in the proliferation of telecommunications (with mobile operators contributing to the high rate of teledensity) in Nigeria, the issue of bandwidth is still a nightmare. Where internet facilities are

available, they are very costly and too slow to do meaningful work [30] Government should provide a framework and support that will make available affordable broadband through mobile telecommunication operators since they are present in almost all communities in the country.

Awareness. The people cannot adopt a system that is familiar to them. For it to be widely accepted and for its adoption to have a soft landing, the people must be educated about what e-health is all about, they need to know the essence of the new system and the benefits accrued to it. Once they are convinced that the benefits outweigh the stress of migrating to the new system, acceptance is guaranteed. The awareness should also include associated legal, ethical and economic issues associated with e-healthcare system. To solve the issue of awareness, massive publicity and jingles should be made in Electronic and social media outlining the benefits of e-health.

Acceptance. According to the Commonwealth Health Ministers Reference Book [31] "Resistance to change has become rooted in certain professional roles – the introduction of ICT in healthcare disrupts traditional structures and hierarchies. Frequently, professionals are unwilling to collaborate in recording and exchanging patient data, with concomitant distrust for off-site data storage and access control". This issue can also be solved by organizing workshops and seminars on the long term benefits of e-health if accepted and implemented.

Data Standardization. The issues here are low definition level of deliverables, conflict in defining minimum data sets for operational management and healthcare decision making, inability to determine the objectives and functionalities desired for applications [31]. Data standardization that allows communication through open access internet-oriented software languages is needed for the automation of processes and services in this wise.

Set-up Cost. The financial outlay of setting up an e-healthcare is huge. This ranges from conception and implementation cost to maintenance services cost to training and re-training. EHR set-up cost includes hardware and software, conversion of existing paper into electronic format and capacity building for the handling of all the processes. In a study by Schmitt and Wofford in 2002, a 280 bed hospital will require US$19m to convert to EHR over a period of seven years [32]. The maintenance cost of repairing and replacement both hardware and software can be huge particularly when newer versions of the software are not compatible with the existing hardware. In University College Hospital (UCH) Ibadan, Nigeria established in 1952 and currently the largest tertiary health facility in the country, having 60 departments, with 850 bed spaces and 163 examination couches [33], using the same parameter will require about US$69million to convert to EHR maintaining the same standard that is obtainable in the US. Government should make available adequate funds for the initial set up and where there is palsy of funds, public-private-participation (PPP) should be explored.

Skill Acquisition. Healthcare practioners and users are very conversant and comfortable with conventional healthcare practices and are most often not ready to change by acquiring new skill to accomplish the same task but the necessary skill must be acquired. The required skills include computer and web technologies, organizational and managerial competences and leadership. This may also lead to loss in revenue arising from the operators learning the new system which may reduce productivity at this time and in turn lead to revenue loss. All the same, skill acquisition is a must. The long term returns will be worth the trouble.

Security and Privacy. The sensitive nature of healthcare information calls for reliable security and privacy. In other words there must be confidentiality; the information must not be prone to intrusion, theft, fraud, unauthorized access and use, and damage. Security and privacy therefore becomes an issue to be solved before people can trust and fully adopt the system. Access to the EHR must be restricted and based on what is required for each personnel working in consonant with a patient. The levels of access should correspond to the job details of each personnel.

Legal and Ethical Issues. The legal and ethical consequences of using e-healthcare applications (EHR, CPOE, CDS and so on) systems which may result in dangerous outcome in certain instances are not well defined. Government should explain and redefine the regulatory agenda for e-healthcare incentive and their purpose. There is need to engage quality and safety assurance technique by system developers so as to avoid clinical hazards and legal liability

Data Storage and System Upgrade. The issue of hardware system breakdown and data loss requires that a secondary databank should be provided to serve as a backup in case of system collapse or upgrade. Extra care must be taken to safe guard loss of clients (patients) data.

Unplanned Aftereffects. e-Healthcare may trigger several unplanned aftereffects, for instance increased medical errors, adverse emotions, overreliance on technology and fluctuations in power [34]. An increase in medical errors can originate if the end users are not trained effectively and this may lead to strong adverse emotions as the operators struggle to use EHR. Every management of health facilities should ensure that even when there is a breakdown in e-healthcare application, the facility should still be running.

5 Proposed Solution

The Nigeria government should setup a strong team to understudy how the system is working in countries like UK and USA and integrate the same in the healthcare sector of Nigeria. Efforts should be made in mapping out strategies for improved power supply which should include more generating plants [hydro, thermal and nuclear] by the government and private sector participation should also be encouraged. Government should provide a framework and support that will make available affordable broadband through mobile telecommunication operators since they are present in

almost all communities in the country. Also provision of adequate funds must be made for the initial set up of e-healthcare system.

Awareness, training and capacity building should be carried out at all levels, highlighting the benefits of ICT in healthcare system. Confidentiality must be guaranteed and the EHR must not be prone to intrusion, theft, fraud, unauthorized access and use, and damage. The levels of access to data should correspond to the job details of each personnel. Secondary databank should be provided to serve as a backup in case of system collapse or upgrade.

Basic data headings used in the analogue system should be agreed upon and transferred to the digital platform, to enable easy and familiar terms and terrain for the operators and the user of the e-healthcare system. Every management of health facilities should ensure that even when there is a breakdown in EHR, the facility will still be running by making provision for backup system that will be deployed in case of any breakdown or power fluctuation.

5.1 Feasibility of the Proposed Solution

The proposed solution is simple and the feasibility in Nigeria is not far-fetched provided the government is committed to its deployment. Once the public health sector adopts the system with support from the government, the private health sector will find it easier investing in the system. All the limitations highlighted as open issues and challenges can be by passed and the system can actually work here and even thrive well. Success is guaranteed if adequate support as received in the telecommunication sector is extended to e-healthcare adaptability in Nigeria.

5.2 Presentation and Explanation of Table

The table below (Table1) summarizes the existing papers reviewed on identified open issues in Nigeria, stating the issues identified in each of the literatures and issues they failed to identify and the last column of the table are the solutions we suggested. Idowu et al, 2008 for example discussed healthcare risk due to insufficient healthcare delivery scheme but did not discussed the extent of the risk and how ICT can used to improve the risk. We suggested that application of ICT in developed countries like UK and USA could be understudied and integrate the same in Nigeria. Premium Times, 2013 raised the issue of power generation and distribution as insufficient to meet the demands of the populace but did not discussed how to resolve the power supply challenge and we have suggested in our paper that efforts should be made in mapping out strategies for improved power supply which should include more generating plants [hydro, thermal and nuclear] by the government and private sector participation should be encouraged. Rodrigues, 2008 mentioned underutilization of ICT in healthcare sector and stated that awareness, training and capacity building should be carried out at all levels, highlighting the benefits of ICT in healthcare system. Other reviews can be found in table1 below.

Table 1. Open Issues in e-Healthcare Delivery System in Nigeria and Proposed Solutions

Existing Work on open issues in Nigeria	Issues discussed	Issues not discussed	Solution proffered in this paper
Idowu et al, (2008) Ref [25]	Health risk as a result of insufficient healthcare delivery scheme	The extent of the possible risk and how ICT can ameliorate the risk	Understudy the application of ICT in healthcare system of developed nations such as UK and integrate same to suit Nigeria
Premium Times (2013) Ref [28]	Power generation and distribution insufficient to meet the demand of the populace	How to resolve the epileptic power supply challenge	Efforts should be made in mapping out strategies for improved power supply which should include more generating plants [hydro, thermal and nuclear] by the government and private sector participation should be encouraged
Roberto, R.J. (2008) Ref [23]	Underutilization of ICT in healthcare system		Awareness, training and capacity building should be carried out at all levels, highlighting the benefits of ICT in healthcare system
Henry Ifeanyi (2013) Ref [30]	Costly and slow Internet facilities to do any meaningful operation		Government should provide a framework and support that will make available affordable broadband through mobile telecommunication operators since they are present in almost all communities in the country
Schmitt and Wofford (2002) Ref [31]	Huge Set up cost of e-Healthcare system	The actual cost required to set-up e-healthcare in Nigeria was not discussed	The government should make available adequate funds for the initial set up [as a correlative estimate for UCH Ibadan will take over $69million, inflation not considered] and where there is palsy of funds, public-private-participation (PPP) should be explored
Idowu et al, (2008) Ref [25]	Data standardization as a result of low definition level of		Basic data headings used in the analogue system should be agreed upon and transferred to the digital platform, to enable easy

Table 1. (*Continued*)

	deliverables, conflict in defining minimum data sets for operational management and healthcare decision making, inability to determine the objectives and functionalities desired for applications		and well familiar terms and terrain for the operators and the user of the e-healthcare system
Campbell, et.al (2006) Ref [32]	EHR may trigger several unplanned aftereffects, for instance increased medical errors, adverse emotions, overreliance on technology and fluctuations in power		Every management of health facilities should ensure that even when there is a breakdown in EHR, the facility will still be running by making provision for backup system that will be deployed in case of any breakdown or power fluctuation

5.3 The Proposed Healthcare Transition in Nigeria

Figure 1 depicts a proposed healthcare transition in Nigeria from the conventional healthcare system (CHS) to e-healthcare system. The arrow from CHS to opens issues indicates that there are basic issues faced by Nigeria hindering its successful transition, the presence of which cannot translate to e-healthcare. The arrow going from conventional healthcare system to e-healthcare enabled delivery system depicts a successful transition in the absence or resolved open issues.

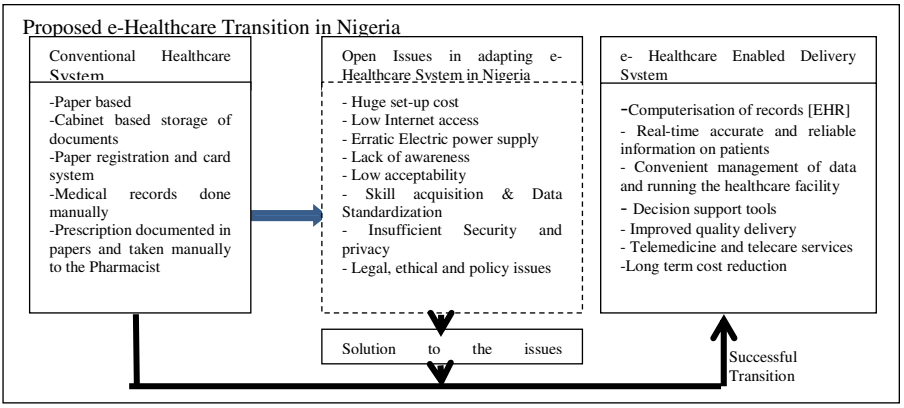


Fig. 1. Proposed e-healthcare Transition in Nigeria

5 Conclusion

The basic concepts about e-healthcare have been discussed in this paper with concentrations on the issues arising from the adoption of the system in the Nigeria context. Issues such as low internet access, erratic electric power supply, lack of awareness, low acceptability, legal and ethical issues and many more were looked into. Some of the solutions given center round the government; to provide framework and support that will make available affordable bandwidth, appropriate health delivery scheme, improvement in power supply by building more generating plants, creating adequate awareness, training and capacity building and many more. Putting the suggested solutions discussed into consideration guarantees the successful adoption of e-healthcare system which then translates to an enhanced healthcare delivery with full benefits of the system.

References

1. Tierney, W.M., McDonald, C.J., Martin, D.K., Rogers, M.P.: Computerized Display of Past Test Results. Ann. Intern. Med. 4, 569–574 (1987)
2. Overhage, J.M., Dexter, P.R., Perkins, S.M.: A Randomized Controlled Trial of Clinical Information Shared from Another Institution. Ann. Emerg. Med. 30(1), 14–23 (2002)
3. Gustafson, D.H., Hawkins, R., Boberg, E., Pingree, S., Serlin, R.E., Graziano, F., Chan, C.L.: Impact of a Patient-Centered, Computer-Based Health Information/Support System. Am. J. Prev. Med. 16(1), 1–9 (1999)
4. Teich, T.M., Wagner, M.M., Mackenzie, C.F.: The Informatics Response in Disaster Terrorism and War. J. Am. Med. Inform. Assoc. 2, 97–104 (2002)
5. Gurstein, M.: Effective Use: A Community Informatics Strategy beyond the Digital Divide. First Monday 8, 12 (2003)
6. United Nations Millennium Declaration. UN General Assembly Resolution A/RES/55/2, http://www.un.org/millennium/declaration/ares552e.htm (accessed October 12, 2014)
7. Pagliari, C., Sloan, D., Gregor, P., Sullivan, F., Detmer, D., Kahan, J.P., Oortwijn, W., MacGillivray, S.: What Is E-Health. Journal of Medical Internet Research 7, 1 (2005)
8. European Commission, A.: Digital Agenda for Europe, European Commission: Brussels, 245 (2010)
9. Radaelli, C.M.: Policy Learning and the Open Method of Coordination in Policy Learning and the New European Economic Governance ESRC. Future of Governance Programme London (2003)
10. Lapointe, L.M., Mignerat, I.V.: The IT Productivity Paradox in Health: A Stakeholder's Perspective. Int. J. Med. Inform. 2, 102–115 (2011)
11. Austin, C., Boxeman, S.: Information System for Healthcare Management, 6th edn. Health Administration Press, Chicago (2008)
12. http://www.searchhealthit.techtarget.com/definition/HITECH-Act (accessed July 12, 2014)
13. Institute of Medicine.: Crossing the Quality Chasm: A New Health System for the 21st Century. National Academies Press, Washington, DC (2001)
14. Anastasium, M.: E-Health: A Bridge to People-Centered Health Care.: Hellenic Society for Quality & Safety in Healthcare & European Society for Quality in Healthcare, Greece (2011)

15. AHRQ: National Healthcare Quality Report, http://archive.ahrq.gov/research/findings/nhqr09/key.html (accessed July 12, 2009)
16. Mekhijian, H.S., Kumar, R.R., Kuehn, L., Bentley, T.D., Treater, P., Thomas, A., Payne, B., Ahmad, A.: Immediate Benefits Realized Following Implementation of Physician Order Entry at an Academic Medical Center. J. Am. Med. Inform. Assoc. 9, 529–539 (2002)
17. Pascal, B.: Medication Errors: Prevention using Information Technology Systems. British Journal of Clinical Pharmacological 67(6), 681–686 (2009)
18. California Health Care Foundation: Computerized Physician Order Entry. Oakland, California (2000), http://quality.chcf.org (accessed October 5, 2014)
19. Kavistha, A.D.: The Most Efficient Health Care System in the World. The Huffington Post (2013)
20. Bates, D.W., Evans, R.S., Murff, H., Stetson, P.D., Pizziferri, L., Hripcsak, G.: Detecting Adverse Events Using Information Technology. JAMIA 10, 115–128 (2003)
21. Kuperman, G.: Improving Response to Critical Laboratory Results with Automation: Results of a Randomized Controlled Trial. JAMIA 5, 2–22 (1999)
22. Douglas, S.B., Susan, G.S., Shinyi, W., Alice, H.C., Margot, B.K.: Use of an Electronic Referral System to Improve the Outpatient Primary Care-Specialty Care Interface. AHRQ 11, 12 (2012)
23. Roberto, R.J.: Compelling Issues for Adoption of e-Health. The Commonwealth Health Ministers Reference Book (2008)
24. Nigerian Census, N.P.C.: Releases 2006 Population Figures. The News online, Village Square (2006)
25. Idowu, P., Cornford, D., Bastin, L.: Health Informatics Deployment in Nigeria. Journal of Health Informatics in Developing Countries 2, 15–23 (2008)
26. David, A.H., David, S.: The Poor State of Health Care Quality in the US: Is Malpractice Liability Part of the Problem or Part of the Solution? Research Paper Series, Paper No 8. University of Texas School of Law (2004)
27. NHS: Information for Health: An Information Strategy for the Modern NHS. Leeds (1998)
28. Premium Times (online): Nigeria's Power Generation Declines to 10-Month Low Say Presidential Task Force, http://m.premiumtimesng.com/news/147010-nigerias-power-generation-declines-10-month-low-says-presidential-task-force.html (accessed August 14, 2014)
29. Claudius, A.A.: Nigerian Electricity Industry: Issues, Challenges and Solutions. Public Lecture Series, vol. 3 (2014)
30. Henry, I.: No Quick Fix to Nigeria's Internet Problem (2013), http://m.news24.com/nigeria/SciTech/News/No-quick-fix-to-Nigerias-internet-problems-20130423 (accessed August 15, 2014)
31. The commonwealth Secretariat: Commonwealth Health Ministers Reference Book 2008. Commonwealth secretariat (2008), http://www.ehealthstrategies.com (accessed August 14, 2014)
32. Schmitt, K.F., Wofford, A.: Financial Analysis Projects Clear Returns from Electronic Medical Records. Health Financ. Manage. 1, 52–57 (2002)
33. University College Hospital, Ibadan: http://uch-ibadan.org.ng/about_uch (accessed September 20, 2014)
34. Campbell, E.M., Sitting, D.F., Ash, J.S.: Types of Unintended Consequences Related to Computerized Provider Order Entry. J. Am Med. Inform. Assoc. PubMed. 5, 547–556 (2006)

Author Index